

LinearRegresssionGenderScopus2minimalRobustLoop.R

MikeT

Wed Jun 12 11:31:53 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,

```

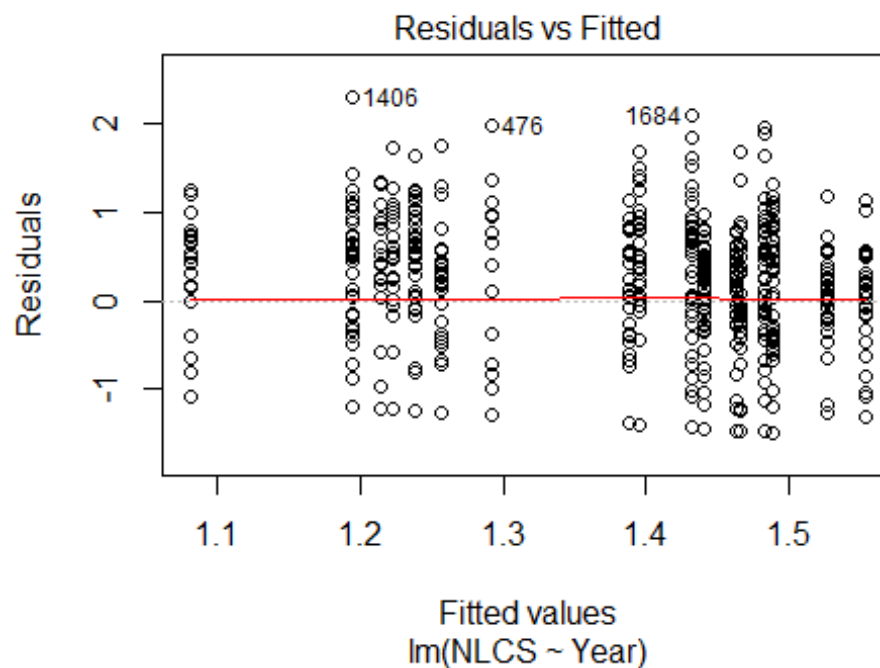
```

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"))
}
}

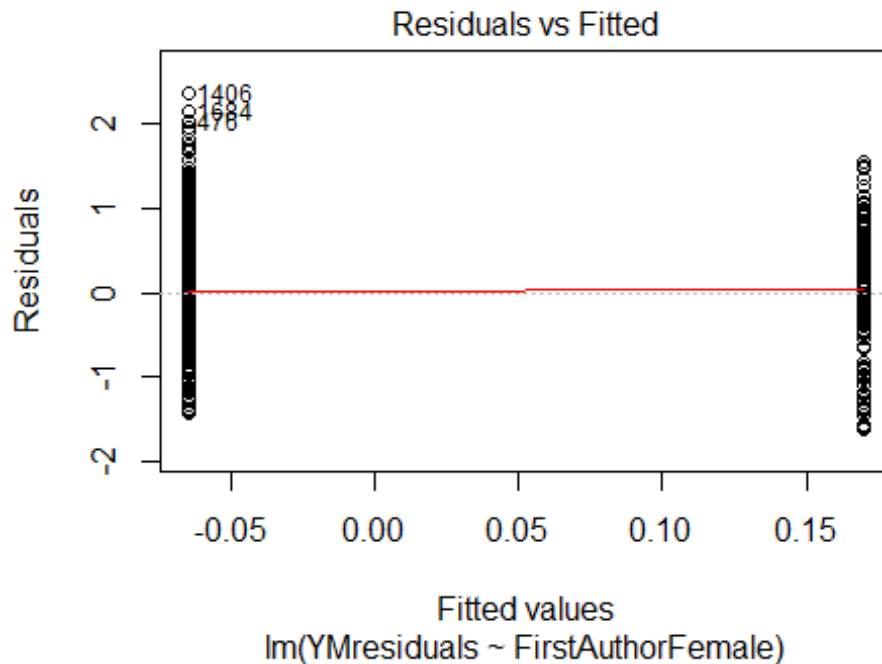
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1000"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 78 59 62 64 51 64 46 53 50 78 62 66 80 84
## 2011 2012
## 100 67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 28 49 31 44 32 24 36 28 36 34 51 35 26 52 53
## 2011 2012
## 49 54
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    26   46   27   42   29   23   33   22   28   30   42   31   22   46   47
## 2011 2012
##    41   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 = 54, df = 16, p-value = 5e-06
```

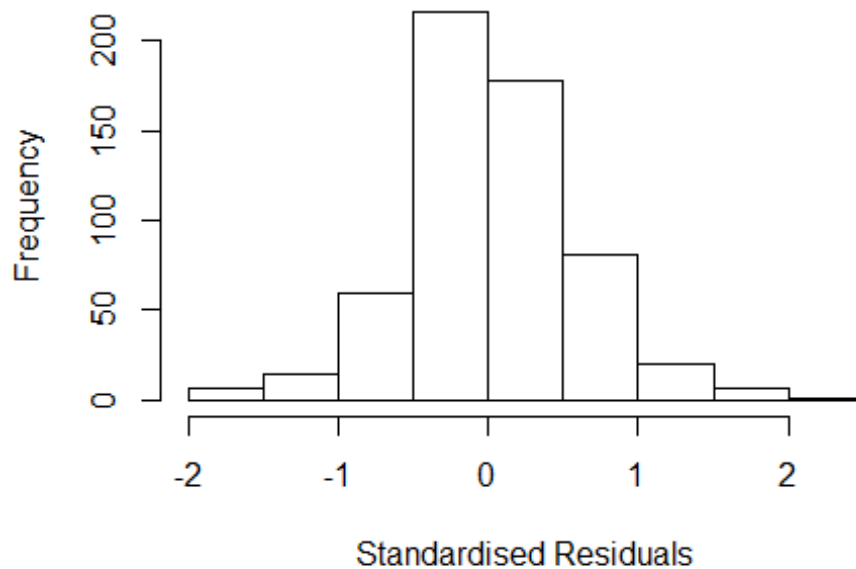


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.1, df = 1, p-value = 0.008
```

```
## [1] "Female first author team size 2018 geometric mean: 4.36508004298606"
## [1] "Male first author team size 2018 geometric mean: 3.70241741766312"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9400, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.13937045759512"
## [1] "Male last author team size 2018 geometric mean: 3.94097720707048"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7200, 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 1.314 1      1.146
## LastAuthorFemale  1.331 1      1.154
## UniqueAuthors     2.907 4      1.143
## Year               3.190 16     1.037
```

Residuals from first and last author and team size



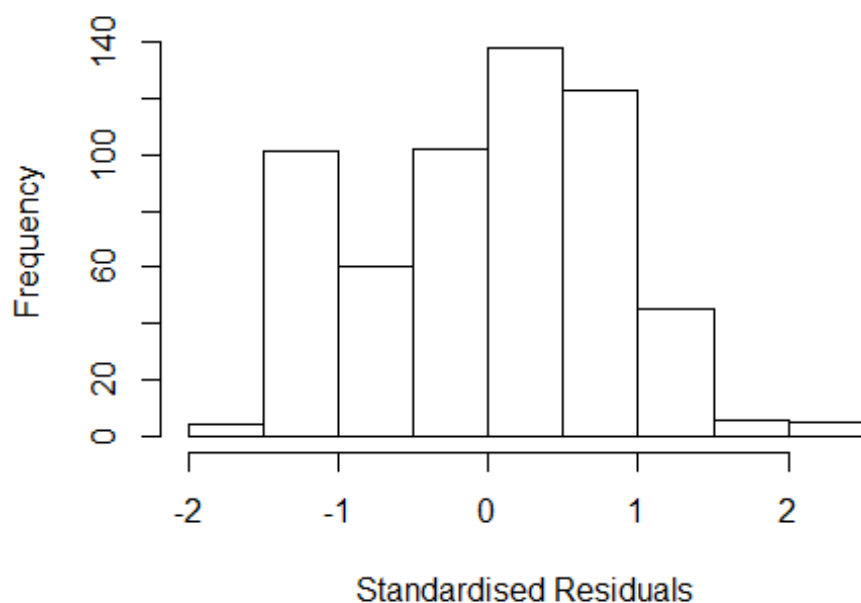
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9246 -0.3843 -0.0168  0.3651  2.2393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.38432    0.14434   2.66   0.008 **
## FirstAuthorFemale1 0.01800    0.05377   0.33   0.738
## LastAuthorFemale1 -0.08831    0.07733  -1.14   0.254
## UniqueAuthors2    1.03965    0.14007   7.42 4.3e-13 ***
## UniqueAuthors3    1.31658    0.10928  12.05 < 2e-16 ***
## UniqueAuthors4    1.51447    0.09691  15.63 < 2e-16 ***
## UniqueAuthors5    1.56564    0.07071  22.14 < 2e-16 ***
## Year1997          0.07623    0.17262   0.44   0.659
## Year1998          0.01096    0.16941   0.06   0.948
## Year1999          0.19868    0.18171   1.09   0.275
```

```

## Year2000      -0.09863    0.15536   -0.63    0.526
## Year2001      0.02421    0.17963    0.13    0.893
## Year2002      0.00628    0.16341    0.04    0.969
## Year2003     -0.12718    0.18374   -0.69    0.489
## Year2004      0.05006    0.15548    0.32    0.748
## Year2005     -0.14207    0.15792   -0.90    0.369
## Year2006     -0.18248    0.14961   -1.22    0.223
## Year2007     -0.25845    0.17962   -1.44    0.151
## Year2008     -0.02050    0.18066   -0.11    0.910
## Year2009      0.08707    0.16169    0.54    0.590
## Year2010     -0.14400    0.16406   -0.88    0.380
## Year2011      0.05445    0.18422    0.30    0.768
## Year2012      0.06297    0.16991    0.37    0.711
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.617, Adjusted R-squared:  0.602
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 160 is an outlier with |weight| = 0 ( < 0.00017);
## 48 weights are ~= 1. The remaining 535 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0415 0.8340 0.9340 0.8620 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          1.71e-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 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.125 1 1.061
## LastAuthorFemale 1.092 1 1.045
## Year 1.226 16 1.006

```

Residuals from first and last author



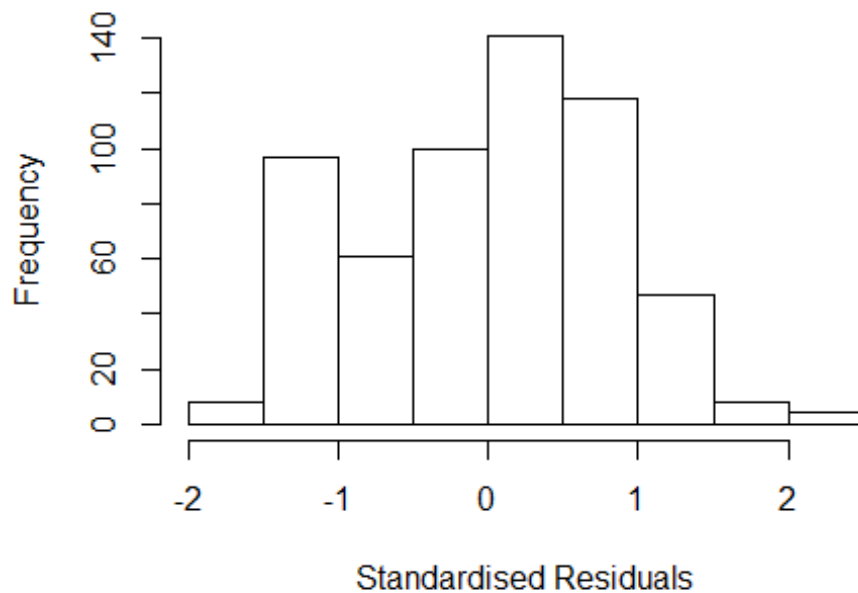
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5640 -0.6491 0.0626 0.6122 2.1743
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0270 0.1753 5.86 7.9e-09 ***
## FirstAuthorFemale1 0.3652 0.0733 4.99 8.2e-07 ***
## LastAuthorFemale1 -0.1499 0.1174 -1.28 0.202
## Year1997 0.1036 0.2392 0.43 0.665
## Year1998 -0.0389 0.2619 -0.15 0.882
## Year1999 0.3155 0.2748 1.15 0.251
## Year2000 0.0313 0.2759 0.11 0.910
## Year2001 0.1594 0.3707 0.43 0.667
## Year2002 0.3895 0.2103 1.85 0.065 .
## Year2003 0.5640 0.2262 2.49 0.013 *
## Year2004 0.4882 0.2101 2.32 0.021 *
## Year2005 0.5008 0.1948 2.57 0.010 *
```

```

## Year2006          0.4138      0.1915      2.16      0.031 *
## Year2007          0.1867      0.2201      0.85      0.397
## Year2008          0.3447      0.2350      1.47      0.143
## Year2009          0.2676      0.2191      1.22      0.222
## Year2010          0.0937      0.2188      0.43      0.669
## Year2011          0.3012      0.2505      1.20      0.230
## Year2012          0.3217      0.2485      1.29      0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.813
## Multiple R-squared:  0.0719, Adjusted R-squared:  0.0423
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 539 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.455  0.852  0.934  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      1.71e-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 1.117 1      1.057
## Year              1.117 16      1.003

```

Residuals from first author

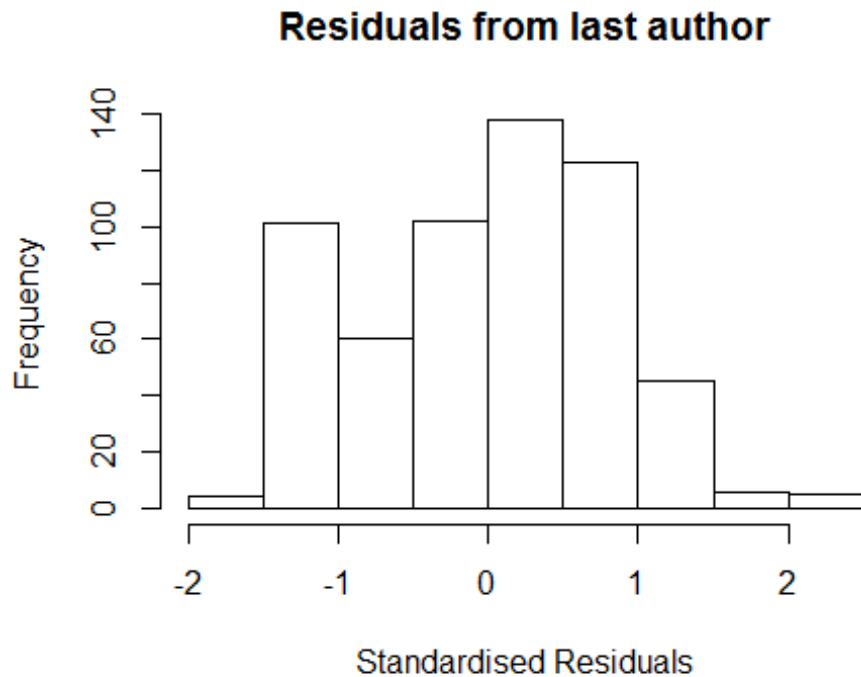


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6814 -0.6531 0.0704 0.6085 2.1867
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02325 0.17491 5.85 8.3e-09 ***
## FirstAuthorFemale1 0.34512 0.07575 4.56 6.4e-06 ***
## Year1997 0.08248 0.23530 0.35 0.726
## Year1998 -0.04483 0.26348 -0.17 0.865
## Year1999 0.30410 0.27339 1.11 0.266
## Year2000 0.00344 0.27312 0.01 0.990
## Year2001 0.14150 0.38995 0.36 0.717
## Year2002 0.38347 0.21180 1.81 0.071 .
## Year2003 0.53162 0.22572 2.36 0.019 *
## Year2004 0.45607 0.20720 2.20 0.028 *
## Year2005 0.49486 0.19479 2.54 0.011 *
## Year2006 0.41053 0.19095 2.15 0.032 *
```

```

## Year2007          0.16940      0.21960      0.77      0.441
## Year2008          0.34398      0.23454      1.47      0.143
## Year2009          0.26312      0.22049      1.19      0.233
## Year2010          0.07997      0.21836      0.37      0.714
## Year2011          0.30170      0.25183      1.20      0.231
## Year2012          0.31306      0.25042      1.25      0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.797
## Multiple R-squared:  0.0702, Adjusted R-squared:  0.0423
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 535 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.432  0.841  0.930  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      1.71e-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"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.102 1      1.050
## Year              1.102 16      1.003

```

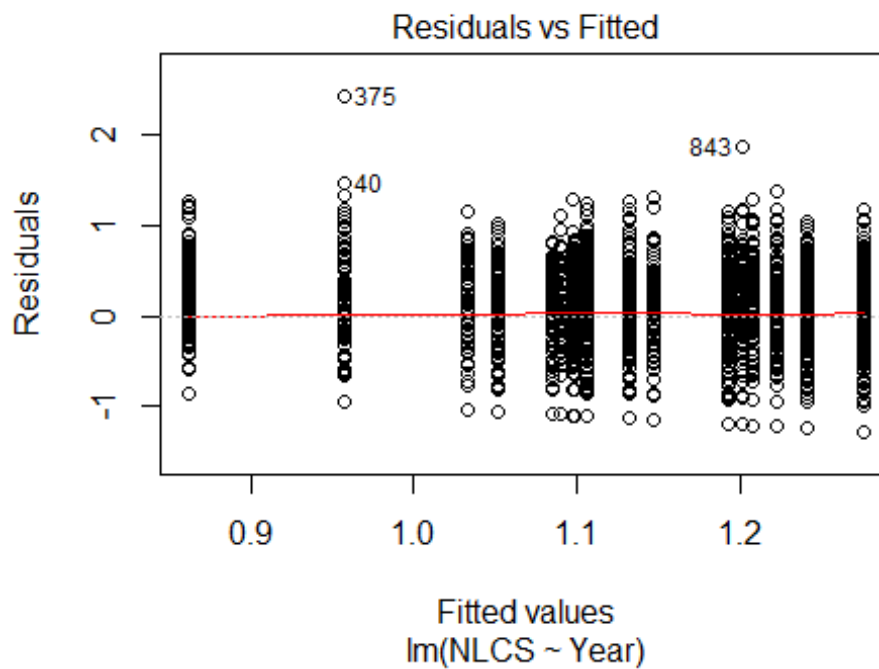
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6349 -0.7047 0.0799 0.6073 2.0422
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0405 0.1791 5.81 1e-08 ***
## LastAuthorFemale1 -0.0574 0.1113 -0.52 0.6059
## Year1997 0.1721 0.2519 0.68 0.4947
## Year1998 0.0507 0.2881 0.18 0.8604
## Year1999 0.3839 0.2794 1.37 0.1701
## Year2000 0.0909 0.2820 0.32 0.7472
## Year2001 0.2256 0.3703 0.61 0.5426
## Year2002 0.4725 0.2104 2.25 0.0251 *
## Year2003 0.5467 0.2274 2.40 0.0165 *
## Year2004 0.5944 0.2059 2.89 0.0040 **
## Year2005 0.5889 0.1950 3.02 0.0026 **
## Year2006 0.5187 0.1937 2.68 0.0076 **
```

```

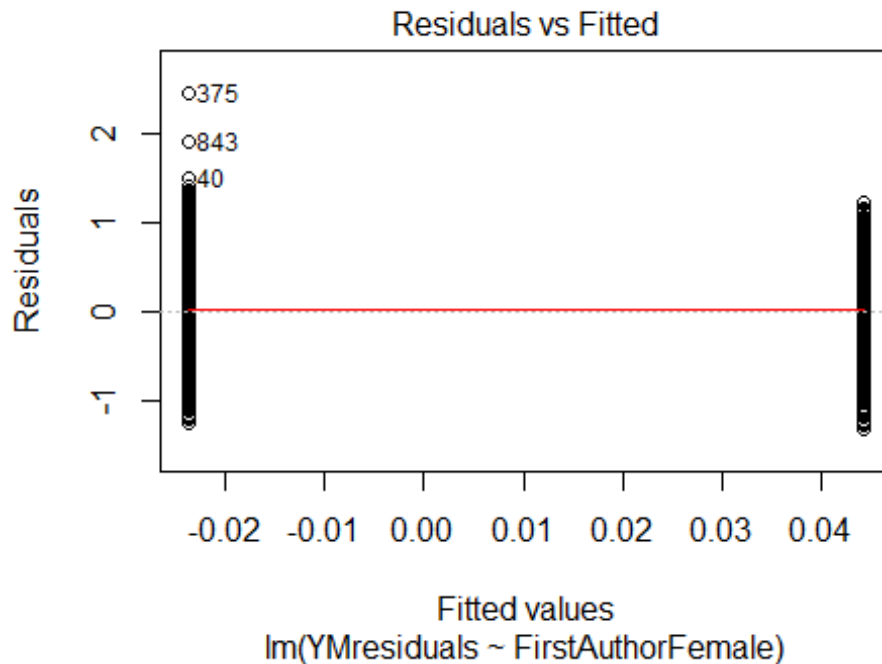
## Year2007          0.2798      0.2247      1.25      0.2135
## Year2008          0.4538      0.2352      1.93      0.0542 .
## Year2009          0.3860      0.2234      1.73      0.0845 .
## Year2010          0.1646      0.2273      0.72      0.4692
## Year2011          0.4198      0.2614      1.61      0.1089
## Year2012          0.4404      0.2459      1.79      0.0739 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.8
## Multiple R-squared:  0.0409, Adjusted R-squared:  0.0121
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 535 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.494  0.832   0.934   0.898   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      1.71e-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: 584"
## [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
##  375  385  404  362  376  350  423  324  296  273  350  337  293  261  287
## 2011 2012
##  506  715
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  106  107  120   98   83   56  177  112   91  100  106  141  147  202  206
## 2011 2012

```

```
## 389 534
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 86 89 110 87 76 42 118 99 79 85 95 129 131 178 188
## 2011 2012
## 343 473
## [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 = 2e-14
```

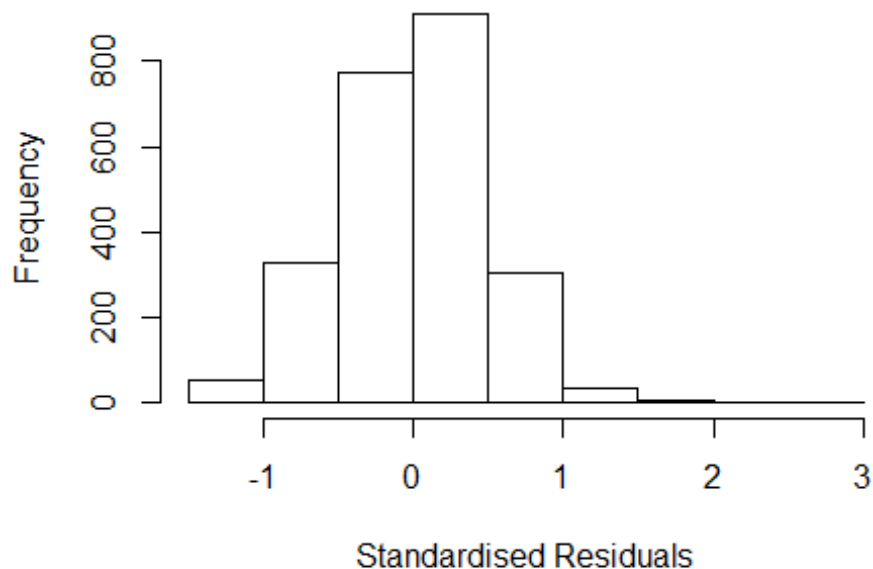


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.36731365231482"
## [1] "Male first author team size 2018 geometric mean: 3.39486103754096"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 25000, p-value = 9e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.18120069272425"
## [1] "Male last author team size 2018 geometric mean: 3.6931345732496"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21000, 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 1.085 1          1.042
## LastAuthorFemale  1.072 1          1.035
## UniqueAuthors    1.533 4          1.055
## Year             1.563 16          1.014
```

Residuals from first and last author and team size



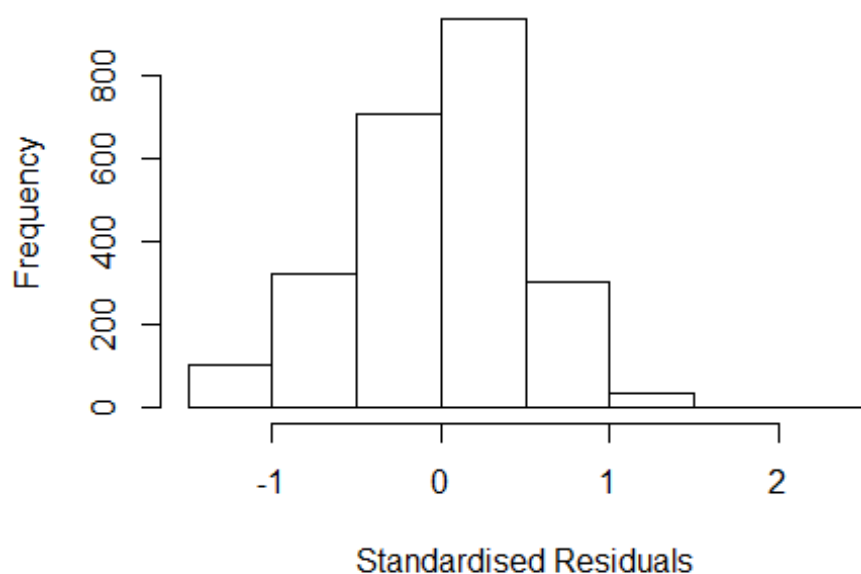
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 375 0039043635 3.387 1996      1100      2      2.615
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4082 -0.3138  0.0227  0.3169  2.6154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77157    0.09077   8.50  <2e-16 ***
## FirstAuthorFemale1 0.01554    0.02100   0.74  0.4595
## LastAuthorFemale1 -0.06538    0.02532  -2.58  0.0099 **
## UniqueAuthors2    0.33138    0.03985   8.32  <2e-16 ***
## UniqueAuthors3    0.45912    0.03818  12.02  <2e-16 ***
## UniqueAuthors4    0.49245    0.04132  11.92  <2e-16 ***
## UniqueAuthors5    0.56154    0.03810  14.74  <2e-16 ***
## Year1997          0.16454    0.10162   1.62  0.1055
## Year1998          0.05802    0.09802   0.59  0.5540
## Year1999          0.10279    0.10233   1.00  0.3152
```

```

## Year2000      -0.06517      0.11297      -0.58      0.5641
## Year2001      0.14414      0.11167      1.29      0.1969
## Year2002     -0.22129      0.10540     -2.10      0.0359 *
## Year2003     -0.02754      0.10358     -0.27      0.7904
## Year2004     -0.01159      0.10027     -0.12      0.9080
## Year2005      0.01136      0.10238      0.11      0.9116
## Year2006      0.01797      0.09962      0.18      0.8568
## Year2007      0.12609      0.09974      1.26      0.2063
## Year2008      0.01325      0.09829      0.13      0.8928
## Year2009     -0.03805      0.09698     -0.39      0.6948
## Year2010      0.03429      0.09437      0.36      0.7164
## Year2011      0.07121      0.09111      0.78      0.4345
## Year2012      0.00604      0.08986      0.07      0.9465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.177, Adjusted R-squared:  0.17
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 78 is an outlier with |weight| = 0 ( < 4.2e-05);
## 185 weights are ~= 1. The remaining 2222 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.187  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           4.15e-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 1.073 1 1.036
## LastAuthorFemale 1.053 1 1.026
## 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
## -1.340 -0.348 0.030 0.339 2.404
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9834 0.0990 9.93 <2e-16 ***
## FirstAuthorFemale1 0.0668 0.0223 3.00 0.0028 **
## LastAuthorFemale1 -0.0662 0.0270 -2.45 0.0143 *
## Year1997 0.2599 0.1117 2.33 0.0201 *
## Year1998 0.1130 0.1095 1.03 0.3023
## Year1999 0.1641 0.1153 1.42 0.1549
## Year2000 0.0318 0.1277 0.25 0.8031
## Year2001 0.1796 0.1239 1.45 0.1475
## Year2002 -0.2054 0.1197 -1.72 0.0864 .
## Year2003 0.0704 0.1167 0.60 0.5463
## Year2004 0.1028 0.1138 0.90 0.3668
## Year2005 0.1382 0.1132 1.22 0.2223
```

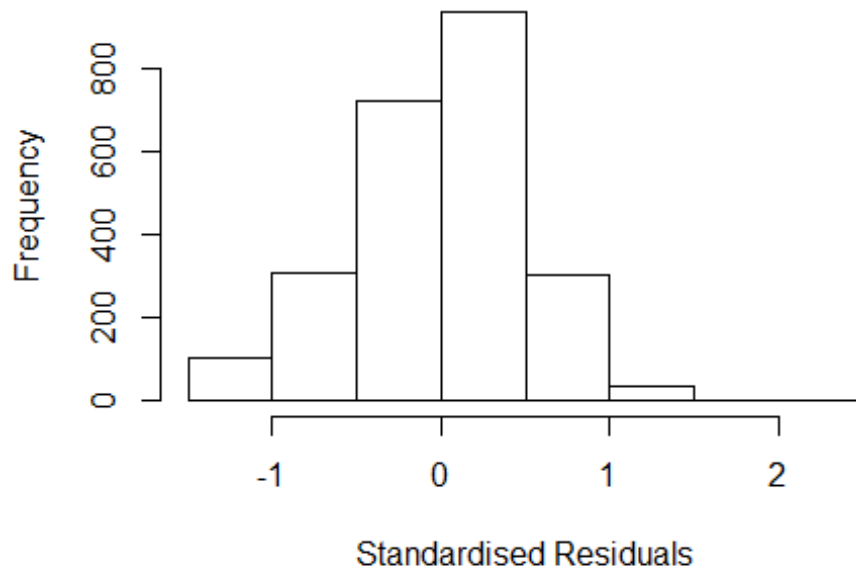


```

## Year2006          0.1811      0.1102      1.64      0.1004
## Year2007          0.2561      0.1100      2.33      0.0199 *
## Year2008          0.1537      0.1099      1.40      0.1620
## Year2009          0.1086      0.1102      0.99      0.3244
## Year2010          0.2178      0.1066      2.04      0.0411 *
## Year2011          0.2896      0.1019      2.84      0.0045 **
## Year2012          0.2503      0.1004      2.49      0.0127 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0561, Adjusted R-squared:  0.049
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## observation 78 is an outlier with |weight| = 0 ( < 4.2e-05);
## 194 weights are ~ = 1. The remaining 2213 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.326  0.864   0.947   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      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 1.055 1          1.027
## Year              1.055 16          1.002

```

Residuals from first author



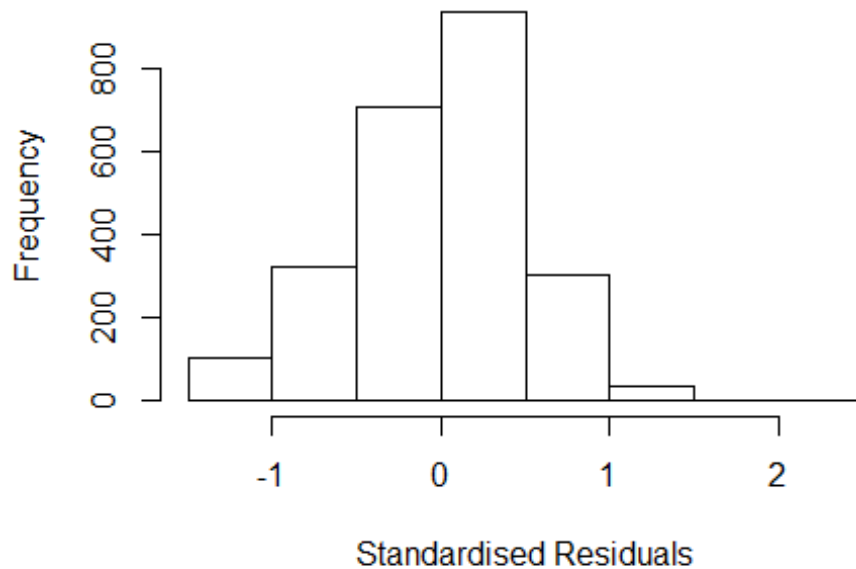
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3194 -0.3434 0.0357 0.3386 2.4146
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9724 0.0999 9.74 <2e-16 ***
## FirstAuthorFemale1 0.0555 0.0223 2.49 0.0127 *
## Year1997 0.2569 0.1126 2.28 0.0227 *
## Year1998 0.1104 0.1105 1.00 0.3183
## Year1999 0.1629 0.1162 1.40 0.1612
## Year2000 0.0342 0.1282 0.27 0.7896
## Year2001 0.1824 0.1248 1.46 0.1440
## Year2002 -0.2067 0.1219 -1.70 0.0902 .
## Year2003 0.0749 0.1172 0.64 0.5227
## Year2004 0.1091 0.1145 0.95 0.3408
## Year2005 0.1445 0.1139 1.27 0.2047
## Year2006 0.1823 0.1116 1.63 0.1025
```

```

## Year2007          0.2570      0.1107      2.32      0.0204 *
## Year2008          0.1575      0.1109      1.42      0.1558
## Year2009          0.1110      0.1111      1.00      0.3181
## Year2010          0.2211      0.1073      2.06      0.0394 *
## Year2011          0.2914      0.1026      2.84      0.0046 **
## Year2012          0.2503      0.1014      2.47      0.0137 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.0538, Adjusted R-squared:  0.0471
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 78 is an outlier with |weight| = 0 ( < 4.2e-05);
## 197 weights are ~= 1. The remaining 2210 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.352  0.861  0.947   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      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 1.029 1          1.015
## Year            1.029 16          1.001

```

Residuals from last author



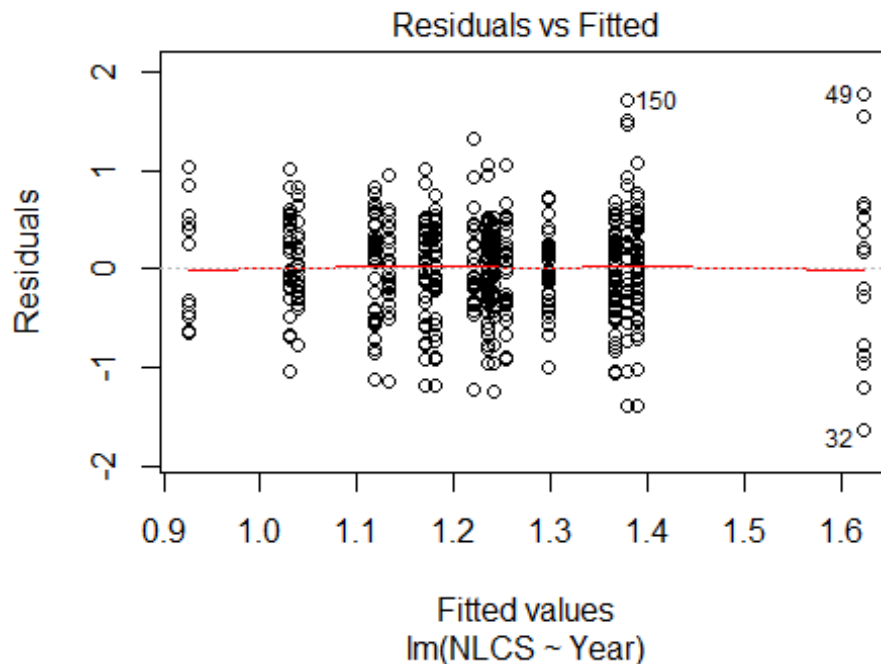
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2978 -0.3483 0.0353 0.3362 2.3907
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9963 0.1011 9.85 <2e-16 ***
## LastAuthorFemale1 -0.0509 0.0265 -1.92 0.0552 .
## Year1997 0.2574 0.1136 2.27 0.0235 *
## Year1998 0.1128 0.1115 1.01 0.3117
## Year1999 0.1658 0.1169 1.42 0.1564
## Year2000 0.0450 0.1287 0.35 0.7267
## Year2001 0.1800 0.1254 1.44 0.1514
## Year2002 -0.2063 0.1218 -1.69 0.0904 .
## Year2003 0.0746 0.1181 0.63 0.5276
## Year2004 0.1133 0.1156 0.98 0.3273
## Year2005 0.1490 0.1143 1.30 0.1925
## Year2006 0.1857 0.1124 1.65 0.0986 .
```

```

## Year2007          0.2626      0.1114      2.36      0.0185 *
## Year2008          0.1628      0.1120      1.45      0.1462
## Year2009          0.1151      0.1119      1.03      0.3040
## Year2010          0.2304      0.1082      2.13      0.0333 *
## Year2011          0.3015      0.1035      2.91      0.0036 **
## Year2012          0.2644      0.1022      2.59      0.0098 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.0465
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 78 is an outlier with |weight| = 0 ( < 4.2e-05);
## 190 weights are ~= 1. The remaining 2217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.339  0.861  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      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: 2408"
## [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
##   52   77   52   43   34   57   55   48   45   55   74   74   64   37   38
## 2011 2012
##   58   61
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   30   28   24   13   26   33   36   32   37   54   59   45   22   31

```

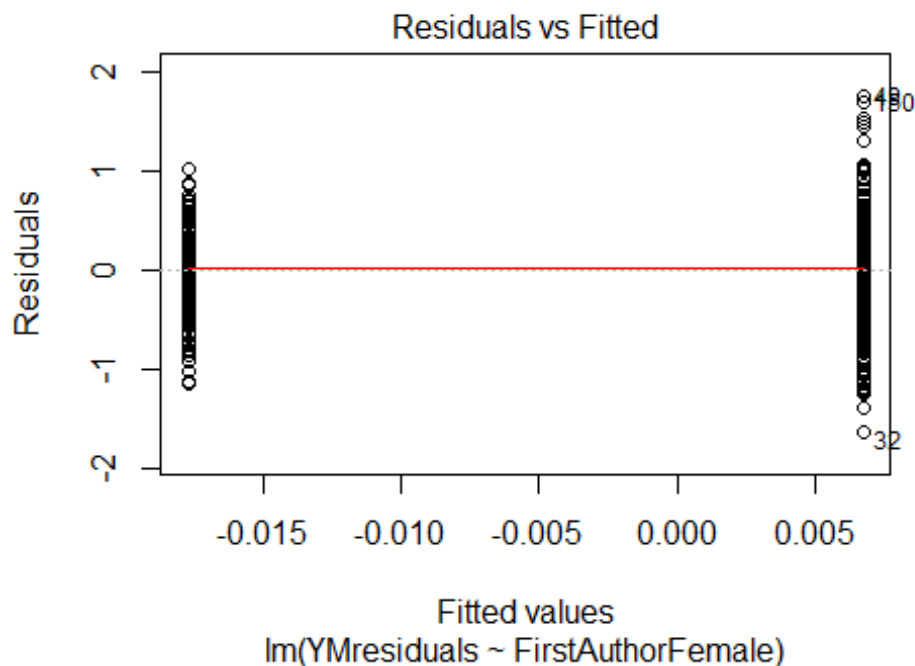
```
## 2011 2012
## 45 48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 26 25 21 12 22 27 29 31 33 42 51 38 21 30
## 2011 2012
## 40 43
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## 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 = 7.7, df = 1, p-value = 0.005
## [1] "Female first author team size 2018 geometric mean: 2.64693141553064"
## [1] "Male first author team size 2018 geometric mean: 2.25314066484579"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

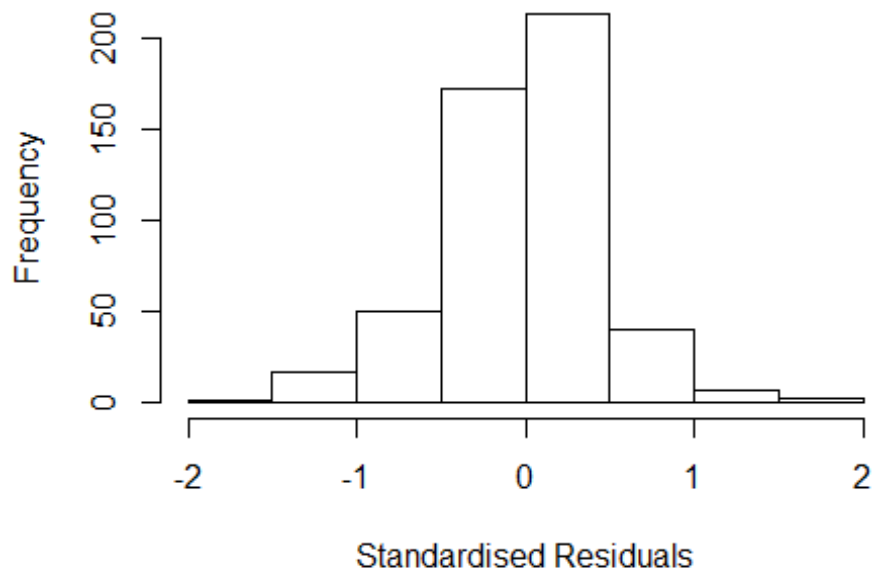
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.06915206929219"
## [1] "Male last author team size 2018 geometric mean: 2.50288630321001"

## 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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.345 1      1.160
## LastAuthorFemale  1.320 1      1.149
## UniqueAuthors    2.406 4      1.116
## Year              3.093 16     1.036
```


Residuals from first and last author and team size



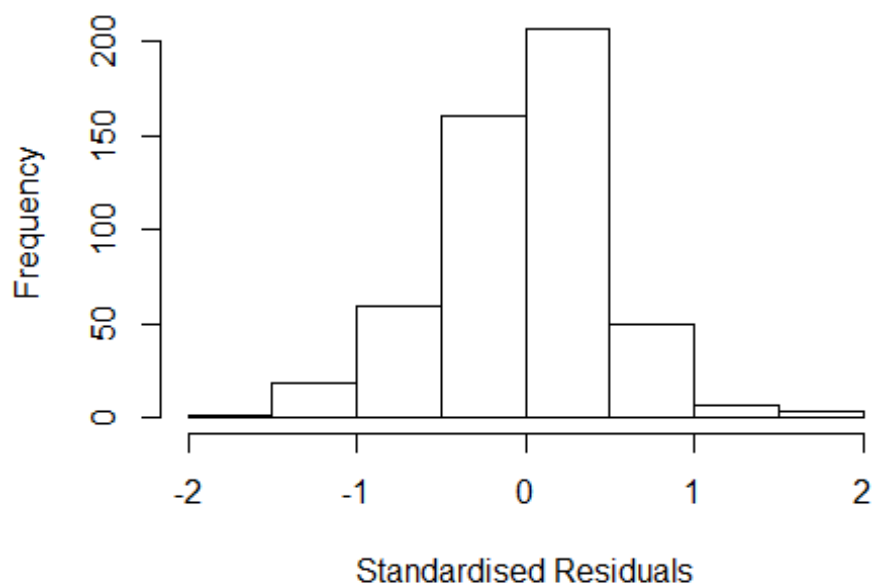
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7658 -0.2687  0.0184  0.2924  1.8425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.5445    0.3617   4.27 2.4e-05 ***
## FirstAuthorFemale1 -0.0645    0.0472  -1.37 0.17271
## LastAuthorFemale1 -0.0829    0.0542  -1.53 0.12679
## UniqueAuthors2     0.2214    0.0610   3.63 0.00031 ***
## UniqueAuthors3     0.2320    0.0619   3.75 0.00020 ***
## UniqueAuthors4     0.3016    0.0726   4.15 3.9e-05 ***
## UniqueAuthors5     0.3467    0.0774   4.48 9.3e-06 ***
## Year1997         -0.3110    0.3895  -0.80 0.42511
## Year1998         -0.5952    0.3732  -1.59 0.11138
## Year1999         -0.6793    0.3608  -1.88 0.06034 .
```

```

## Year2000          -0.9033      0.3901    -2.32  0.02101 *
## Year2001          -0.5258      0.3666    -1.43  0.15212
## Year2002          -0.5191      0.3634    -1.43  0.15385
## Year2003          -0.5588      0.3647    -1.53  0.12615
## Year2004          -0.5069      0.3632    -1.40  0.16353
## Year2005          -0.4688      0.3627    -1.29  0.19685
## Year2006          -0.4634      0.3578    -1.30  0.19589
## Year2007          -0.4575      0.3566    -1.28  0.20010
## Year2008          -0.3646      0.3542    -1.03  0.30378
## Year2009          -0.3146      0.3691    -0.85  0.39445
## Year2010          -0.4262      0.3627    -1.18  0.24055
## Year2011          -0.3234      0.3558    -0.91  0.36385
## Year2012          -0.2330      0.3618    -0.64  0.51982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.161, Adjusted R-squared:  0.123
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 461 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0127 0.8730 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          1.98e-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 1.220 1          1.104
## LastAuthorFemale 1.283 1          1.132
## Year          1.483 16          1.012

```

Residuals from first and last author



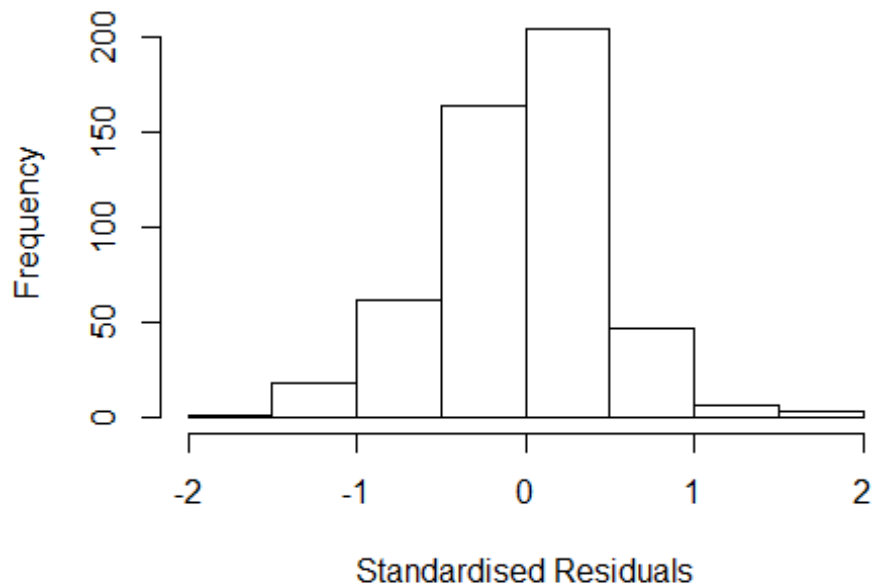
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7281 -0.2885 0.0134 0.3129 1.6589
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7281 0.5287 3.27 0.0012 **
## FirstAuthorFemale1 -0.0179 0.0472 -0.38 0.7050
## LastAuthorFemale1 -0.0897 0.0566 -1.58 0.1137
## Year1997 -0.4297 0.5495 -0.78 0.4346
## Year1998 -0.6164 0.5384 -1.14 0.2528
## Year1999 -0.7322 0.5313 -1.38 0.1688
## Year2000 -0.9094 0.5583 -1.63 0.1040
## Year2001 -0.5811 0.5349 -1.09 0.2778
## Year2002 -0.5761 0.5336 -1.08 0.2808
## Year2003 -0.5784 0.5357 -1.08 0.2808
## Year2004 -0.5523 0.5315 -1.04 0.2993
## Year2005 -0.4923 0.5296 -0.93 0.3530
```

```

## Year2006          -0.4932      0.5275   -0.93   0.3503
## Year2007          -0.4419      0.5270   -0.84   0.4021
## Year2008          -0.3761      0.5254   -0.72   0.4745
## Year2009          -0.2938      0.5338   -0.55   0.5823
## Year2010          -0.4353      0.5309   -0.82   0.4126
## Year2011          -0.2938      0.5257   -0.56   0.5765
## Year2012          -0.2528      0.5305   -0.48   0.6339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.107, Adjusted R-squared:  0.0736
## Convergence in 42 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 463 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0601 0.8580 0.9450 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      1.98e-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 1.191 1      1.091
## Year              1.191 16      1.005

```

Residuals from first author



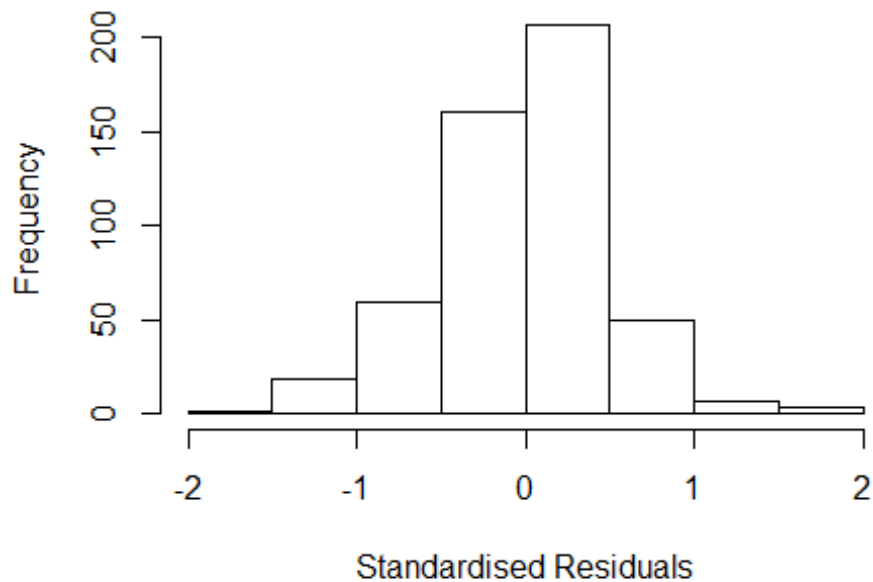
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6988 -0.2852 0.0192 0.3080 1.6882
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6988 0.5927 2.87 0.0043 **
## FirstAuthorFemale1 -0.0315 0.0476 -0.66 0.5080
## Year1997 -0.4290 0.6141 -0.70 0.4851
## Year1998 -0.5941 0.6008 -0.99 0.3232
## Year1999 -0.7079 0.5948 -1.19 0.2346
## Year2000 -0.8956 0.6215 -1.44 0.1503
## Year2001 -0.5524 0.5983 -0.92 0.3563
## Year2002 -0.5645 0.5983 -0.94 0.3459
## Year2003 -0.5532 0.5985 -0.92 0.3558
## Year2004 -0.5279 0.5956 -0.89 0.3759
## Year2005 -0.4787 0.5938 -0.81 0.4205
## Year2006 -0.4883 0.5925 -0.82 0.4103
```

```

## Year2007          -0.4255      0.5913   -0.72   0.4721
## Year2008          -0.3644      0.5904   -0.62   0.5374
## Year2009          -0.2597      0.5965   -0.44   0.6635
## Year2010          -0.4426      0.5956   -0.74   0.4578
## Year2011          -0.2825      0.5906   -0.48   0.6326
## Year2012          -0.2387      0.5947   -0.40   0.6883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.1,   Adjusted R-squared:  0.0685
## Convergence in 46 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 467 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0812 0.8650 0.9470 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      1.98e-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 1.266 1          1.125
## Year            1.266 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
## -1.7171 -0.2843 0.0152 0.3171 1.6699
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7171 0.5224 3.29 0.0011 **
## LastAuthorFemale1 -0.0935 0.0564 -1.66 0.0982 .
## Year1997 -0.4225 0.5455 -0.77 0.4390
## Year1998 -0.6092 0.5345 -1.14 0.2550
## Year1999 -0.7267 0.5284 -1.38 0.1697
## Year2000 -0.9007 0.5533 -1.63 0.1042
## Year2001 -0.5746 0.5315 -1.08 0.2802
## Year2002 -0.5676 0.5291 -1.07 0.2840
## Year2003 -0.5698 0.5308 -1.07 0.2837
## Year2004 -0.5441 0.5270 -1.03 0.3023
## Year2005 -0.4856 0.5265 -0.92 0.3568
## Year2006 -0.4866 0.5240 -0.93 0.3536
```

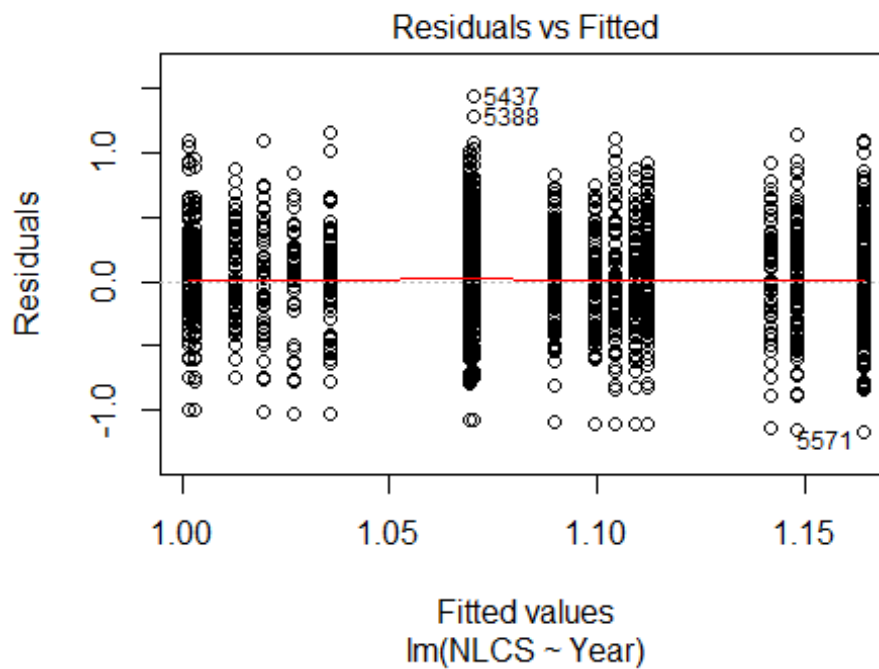
```

## Year2007          -0.4348      0.5230   -0.83    0.4062
## Year2008          -0.3717      0.5230   -0.71    0.4776
## Year2009          -0.2881      0.5306   -0.54    0.5874
## Year2010          -0.4280      0.5272   -0.81    0.4173
## Year2011          -0.2894      0.5235   -0.55    0.5806
## Year2012          -0.2449      0.5264   -0.47    0.6420
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0749
## Convergence in 42 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 461 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0627 0.8570 0.9460 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      1.98e-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: 504"
## [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
## 282 265 272 244 224 213 265 257 214 217 301 301 288 365 395
## 2011 2012
## 353 384
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 46 74 74 66 51 139 110 90 99 141 154 125 166 180
## 2011 2012

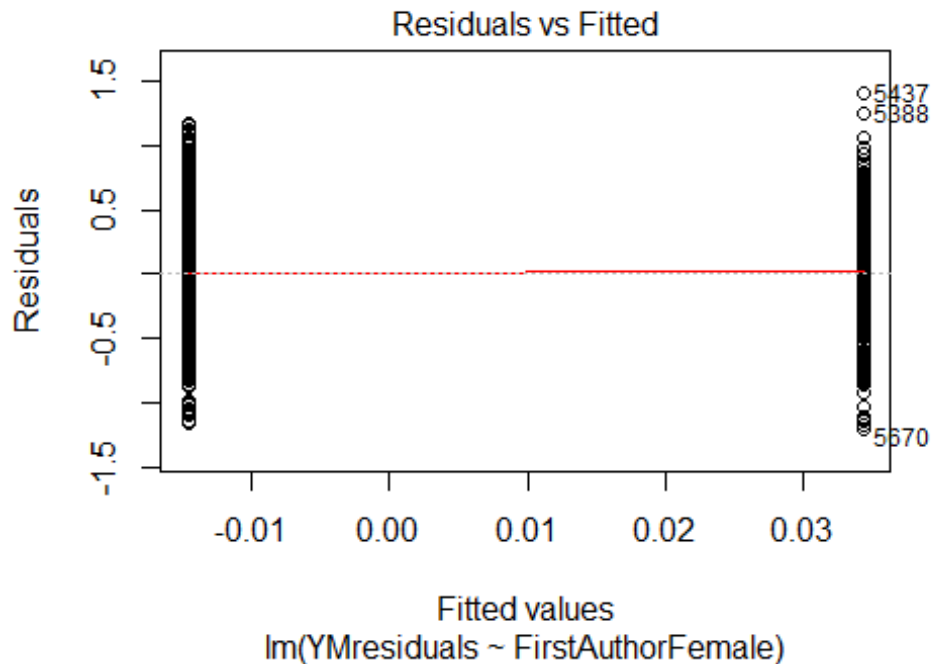
```



```
## 178 178
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 36 71 67 64 42 113 104 76 91 129 136 103 144 157
## 2011 2012
## 154 163
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.009
```

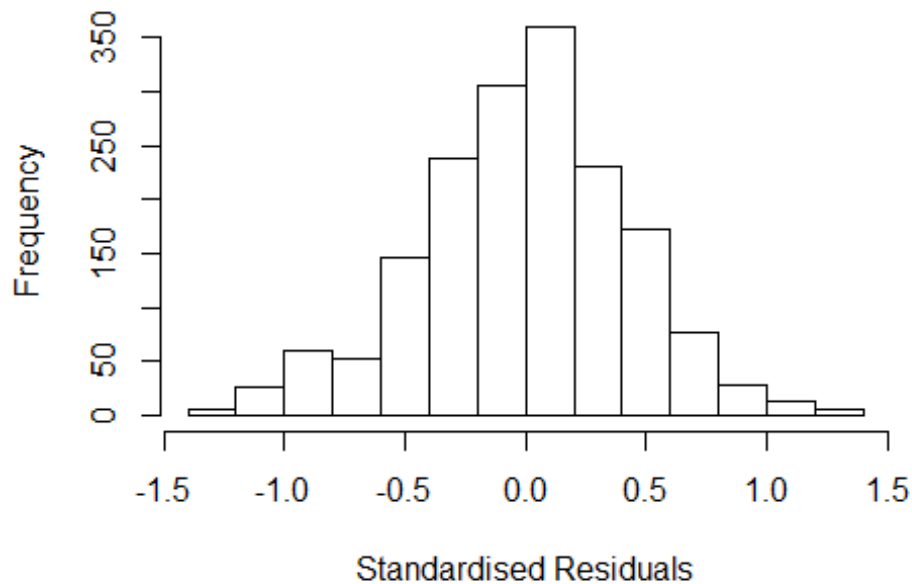


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.73, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 4.158558217014"
## [1] "Male first author team size 2018 geometric mean: 3.46062842081641"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.5423652204211"
## [1] "Male last author team size 2018 geometric mean: 3.51146349543169"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, 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 1.066 1      1.033
## LastAuthorFemale  1.076 1      1.037
## UniqueAuthors    1.291 4      1.032
## Year             1.336 16      1.009
```

Residuals from first and last author and team size



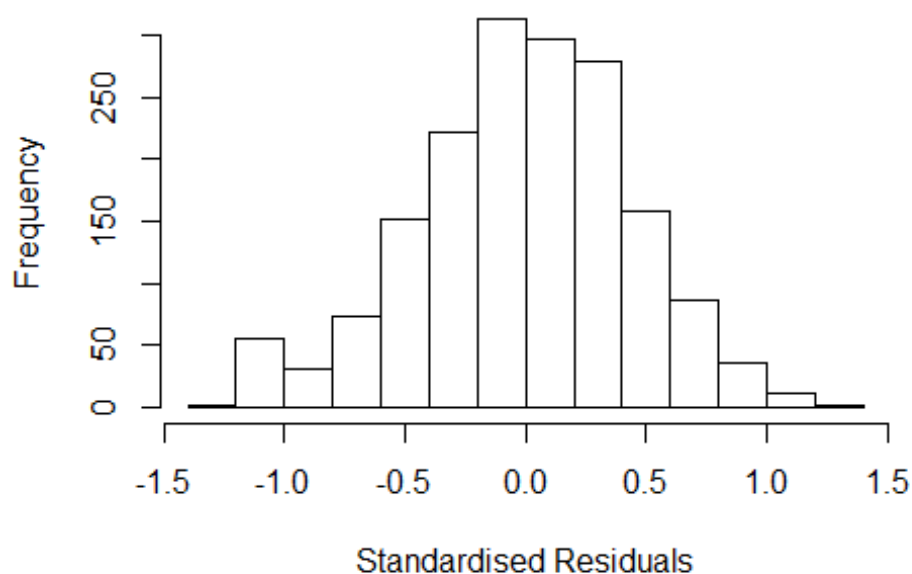
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2854 -0.2743 0.0197 0.2658 1.2856
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92942 0.06063 15.33 < 2e-16 ***
## FirstAuthorFemale1 0.00564 0.02320 0.24 0.808
## LastAuthorFemale1 -0.01711 0.02624 -0.65 0.514
## UniqueAuthors2 0.18314 0.03745 4.89 1.1e-06 ***
## UniqueAuthors3 0.25030 0.03778 6.62 4.7e-11 ***
## UniqueAuthors4 0.30296 0.03938 7.69 2.4e-14 ***
## UniqueAuthors5 0.42669 0.03993 10.69 < 2e-16 ***
## Year1997 -0.09807 0.08112 -1.21 0.227
## Year1998 -0.06727 0.07226 -0.93 0.352
## Year1999 -0.07646 0.06929 -1.10 0.270
```

```

## Year2000      0.03111      0.07995      0.39      0.697
## Year2001     -0.10520      0.08427     -1.25      0.212
## Year2002     -0.15912      0.06590     -2.41      0.016 *
## Year2003     -0.15093      0.06809     -2.22      0.027 *
## Year2004     -0.06081      0.06837     -0.89      0.374
## Year2005     -0.02423      0.06703     -0.36      0.718
## Year2006      0.00558      0.06542      0.09      0.932
## Year2007     -0.07554      0.06611     -1.14      0.253
## Year2008     -0.06989      0.07168     -0.97      0.330
## Year2009     -0.11232      0.06951     -1.62      0.106
## Year2010     -0.10166      0.06827     -1.49      0.137
## Year2011     -0.07633      0.06882     -1.11      0.268
## Year2012     -0.03822      0.06891     -0.55      0.579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0962, Adjusted R-squared:  0.0845
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1573 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.311  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      5.82e-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 1.049 1      1.024
## LastAuthorFemale  1.052 1      1.026
## Year              1.074 16      1.002

```

Residuals from first and last author



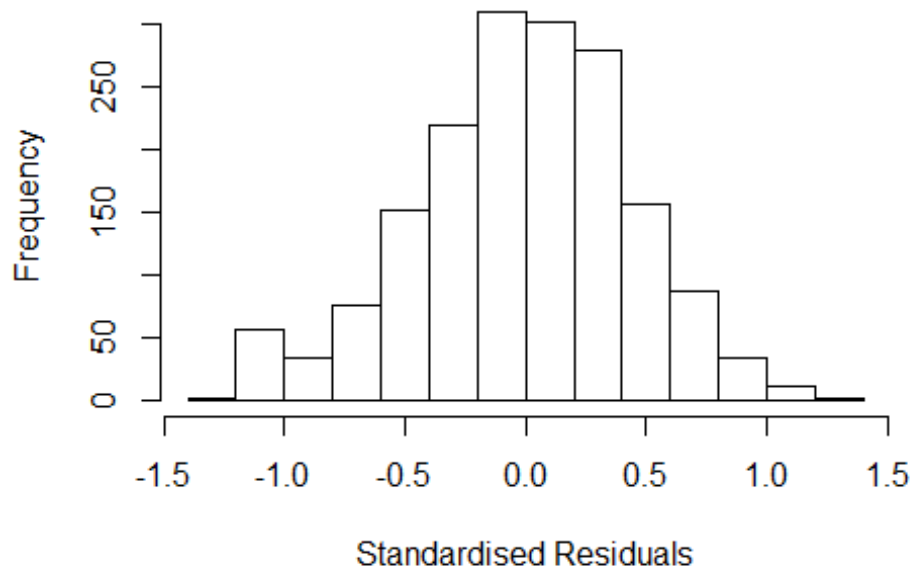
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21117 -0.29222  0.00606  0.29563  1.38026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.12065    0.05823   19.25  <2e-16 ***
## FirstAuthorFemale1  0.04160    0.02399    1.73   0.083 .
## LastAuthorFemale1 -0.02901    0.02749   -1.06   0.291
## Year1997          -0.07484    0.08988   -0.83   0.405
## Year1998          -0.07604    0.07608   -1.00   0.318
## Year1999          -0.12901    0.07355   -1.75   0.080 .
## Year2000           0.04892    0.08379    0.58   0.559
## Year2001          -0.07698    0.08865   -0.87   0.385
## Year2002          -0.11621    0.06851   -1.70   0.090 .
## Year2003          -0.13426    0.07068   -1.90   0.058 .
## Year2004          -0.02424    0.07385   -0.33   0.743
## Year2005          -0.00499    0.07208   -0.07   0.945
```

```

## Year2006          0.04405    0.06912    0.64    0.524
## Year2007          -0.03033    0.06912   -0.44    0.661
## Year2008          -0.02937    0.07498   -0.39    0.695
## Year2009          -0.06676    0.07259   -0.92    0.358
## Year2010          -0.04651    0.07050   -0.66    0.509
## Year2011          -0.03551    0.07192   -0.49    0.622
## Year2012          0.04560    0.07087    0.64    0.520
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.00726
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1575 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.294  0.872  0.949  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      5.82e-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 1.035 1      1.017
## Year              1.035 16      1.001

```

Residuals from first author



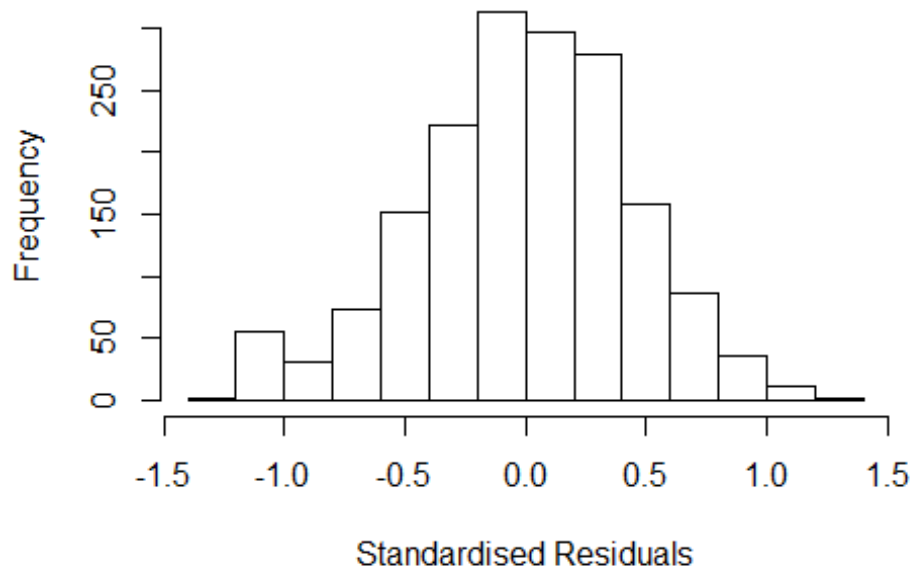
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20026 -0.29199  0.00761  0.29558  1.38852
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11741    0.05819   19.20  <2e-16 ***
## FirstAuthorFemale1 0.03819    0.02390    1.60   0.110
## Year1997      -0.07635    0.08983   -0.85   0.395
## Year1998      -0.07851    0.07596   -1.03   0.301
## Year1999      -0.13067    0.07359   -1.78   0.076 .
## Year2000       0.04466    0.08375    0.53   0.594
## Year2001      -0.07817    0.08842   -0.88   0.377
## Year2002      -0.11737    0.06847   -1.71   0.087 .
## Year2003      -0.13355    0.07069   -1.89   0.059 .
## Year2004      -0.02701    0.07378   -0.37   0.714
## Year2005      -0.00352    0.07201   -0.05   0.961
## Year2006       0.04285    0.06908    0.62   0.535
```

```

## Year2007          -0.03155    0.06920   -0.46    0.648
## Year2008          -0.03040    0.07500   -0.41    0.685
## Year2009          -0.07061    0.07230   -0.98    0.329
## Year2010          -0.04833    0.07060   -0.68    0.494
## Year2011          -0.03713    0.07181   -0.52    0.605
## Year2012           0.04361    0.07080    0.62    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0171, Adjusted R-squared:  0.00725
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1572 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.288  0.870   0.949   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      5.82e-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 1.038 1          1.019
## Year              1.038 16          1.001

```


Residuals from last author



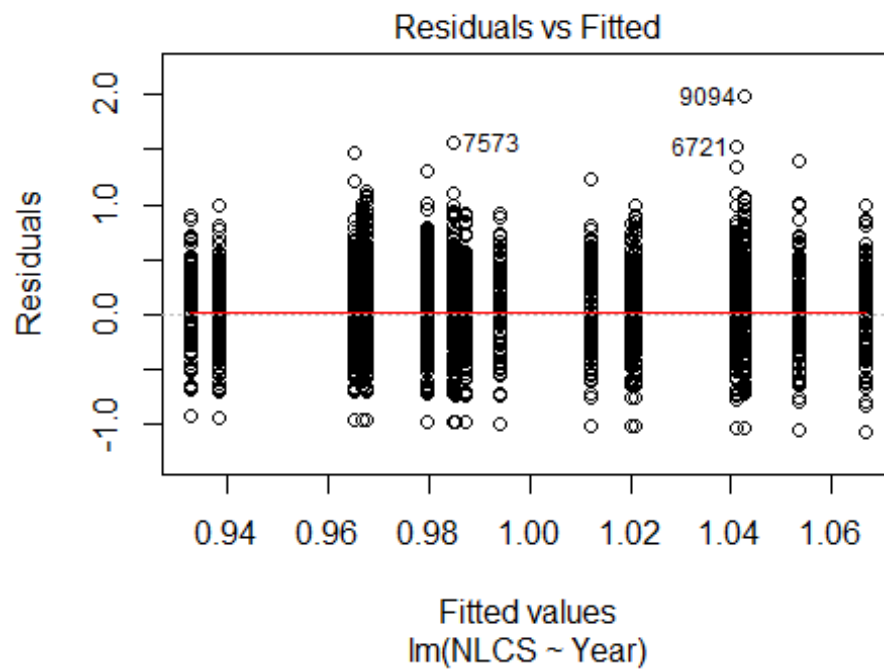
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18281 -0.29139 0.00542 0.29720 1.41034
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12887 0.05840 19.33 <2e-16 ***
## LastAuthorFemale1 -0.02251 0.02713 -0.83 0.407
## Year1997 -0.07704 0.09017 -0.85 0.393
## Year1998 -0.07426 0.07613 -0.98 0.329
## Year1999 -0.12687 0.07410 -1.71 0.087 .
## Year2000 0.05395 0.08381 0.64 0.520
## Year2001 -0.07918 0.08817 -0.90 0.369
## Year2002 -0.11662 0.06887 -1.69 0.091 .
## Year2003 -0.12755 0.07077 -1.80 0.072 .
## Year2004 -0.02055 0.07416 -0.28 0.782
## Year2005 -0.00384 0.07219 -0.05 0.958
## Year2006 0.04697 0.06932 0.68 0.498
```

```

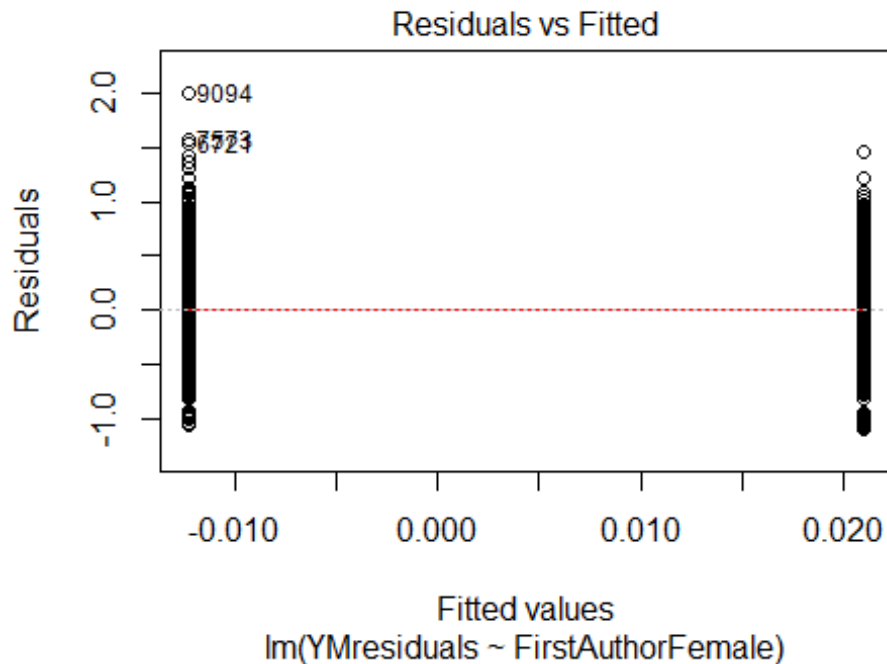
## Year2007          -0.02563      0.06927    -0.37      0.711
## Year2008          -0.02163      0.07511    -0.29      0.773
## Year2009          -0.06437      0.07278    -0.88      0.377
## Year2010          -0.04379      0.07091    -0.62      0.537
## Year2011          -0.03221      0.07220    -0.45      0.656
## Year2012           0.04896      0.07136      0.69      0.493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00613
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1582 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.273  0.874  0.949  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      5.82e-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: 1718"
## [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
##  381  384  351  321  571  388  375  357  336  369  427  413  485  584  651
## 2011 2012
##  571  555
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  146  145  148  144  112  116  183  185  181  214  257  262  316  313  330
## 2011 2012

```

```
## 353 324
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 142 136 136 134 104 104 158 164 162 199 227 238 289 284 293
## 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 = 72, df = 16, p-value = 5e-09
```

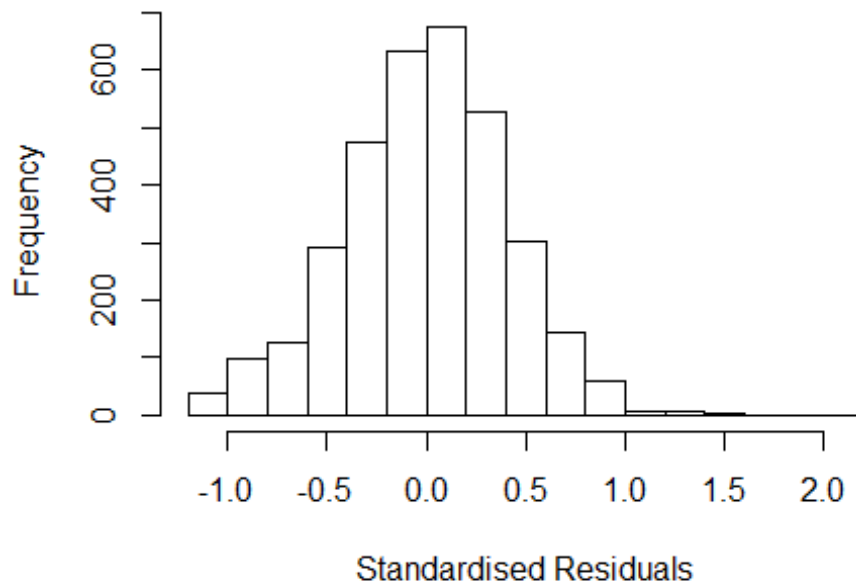


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.1, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.58347883034116"
## [1] "Male first author team size 2018 geometric mean: 3.11683832690531"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.06207025275204"
## [1] "Male last author team size 2018 geometric mean: 3.50411428488961"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.134 1      1.065
## LastAuthorFemale  1.076 1      1.038
## UniqueAuthors    1.225 4      1.026
## Year             1.250 16      1.007
```

Residuals from first and last author and team size



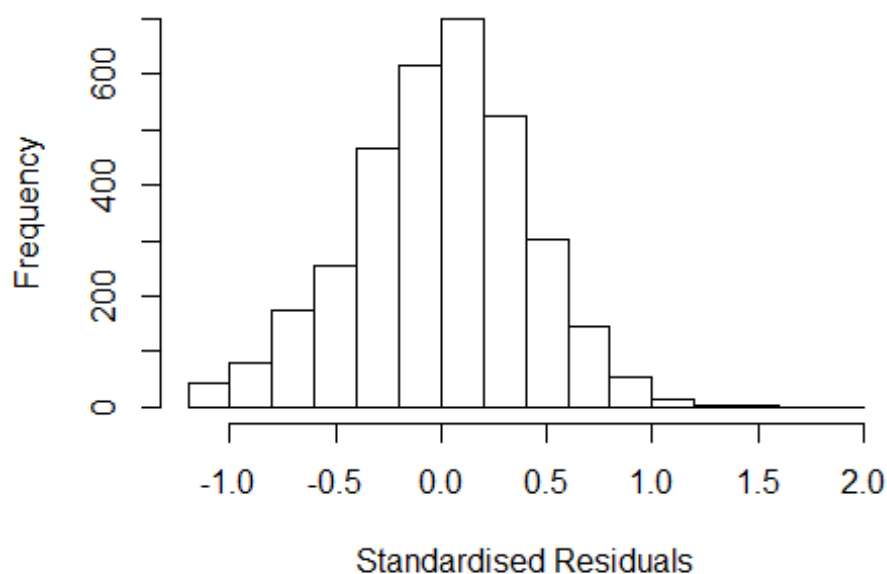
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1380 -0.2638  0.0096  0.2711  2.1130
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.98656    0.03728   26.46 < 2e-16 ***
## FirstAuthorFemale1 0.01869    0.01565    1.19  0.23240
## LastAuthorFemale1 -0.01056    0.01773   -0.60  0.55142
## UniqueAuthors2    0.11788    0.01978    5.96  2.8e-09 ***
## UniqueAuthors3    0.16186    0.02126    7.61  3.5e-14 ***
## UniqueAuthors4    0.19276    0.02717    7.09  1.6e-12 ***
## UniqueAuthors5    0.24177    0.02892    8.36 < 2e-16 ***
## Year1997         -0.08311    0.05221   -1.59  0.11151
## Year1998         -0.09771    0.04780   -2.04  0.04101 *
## Year1999         -0.12754    0.04890   -2.61  0.00914 **
```

```

## Year2000      -0.00811    0.05678   -0.14  0.88643
## Year2001      -0.06000    0.05848   -1.03  0.30495
## Year2002      -0.05290    0.04553   -1.16  0.24539
## Year2003      -0.15144    0.04680   -3.24  0.00122 **
## Year2004      -0.08766    0.04694   -1.87  0.06193 .
## Year2005      -0.12288    0.04520   -2.72  0.00658 **
## Year2006      -0.16331    0.04732   -3.45  0.00056 ***
## Year2007      -0.16118    0.04538   -3.55  0.00039 ***
## Year2008      -0.13210    0.04525   -2.92  0.00353 **
## Year2009      -0.07703    0.04524   -1.70  0.08873 .
## Year2010      -0.14589    0.04533   -3.22  0.00130 **
## Year2011      -0.16081    0.04588   -3.51  0.00046 ***
## Year2012      -0.07159    0.04619   -1.55  0.12125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0459, Adjusted R-squared:  0.0397
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 3292 is an outlier with |weight| = 0 ( < 3e-05);
## 268 weights are ~= 1. The remaining 3118 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.041  0.870  0.951  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           2.95e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "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 1.068 1 1.033
## Year 1.069 16 1.002

```

Residuals from first and last author



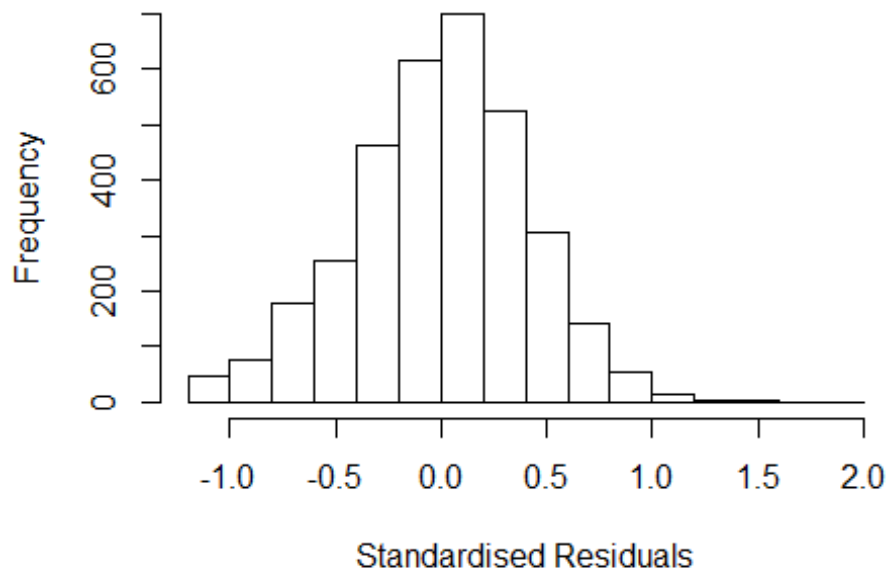
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1174 -0.2687 0.0167 0.2731 1.9757
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05990 0.03753 28.24 <2e-16 ***
## FirstAuthorFemale1 0.04386 0.01558 2.81 0.0049 **
## LastAuthorFemale1 -0.02274 0.01781 -1.28 0.2016
## Year1997 -0.05552 0.05253 -1.06 0.2906
## Year1998 -0.09320 0.04846 -1.92 0.0546 .
## Year1999 -0.12106 0.04942 -2.45 0.0144 *
## Year2000 0.01362 0.05628 0.24 0.8088
## Year2001 -0.05481 0.05775 -0.95 0.3426
## Year2002 -0.02219 0.04526 -0.49 0.6240
## Year2003 -0.11646 0.04685 -2.49 0.0130 *
## Year2004 -0.06475 0.04671 -1.39 0.1657
## Year2005 -0.08550 0.04584 -1.87 0.0622 .
```

```

## Year2006          -0.12092      0.04723    -2.56    0.0105 *
## Year2007          -0.11575      0.04599    -2.52    0.0119 *
## Year2008          -0.08761      0.04596    -1.91    0.0567 .
## Year2009          -0.03348      0.04577    -0.73    0.4645
## Year2010          -0.09738      0.04603    -2.12    0.0345 *
## Year2011          -0.11326      0.04642    -2.44    0.0148 *
## Year2012          -0.00759      0.04647    -0.16    0.8702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.00767
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 3292 is an outlier with |weight| = 0 ( < 3e-05);
## 296 weights are ~= 1. The remaining 3090 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0846 0.8700 0.9500 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          2.95e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.05 1          1.025
## Year              1.05 16          1.002

```


Residuals from first author



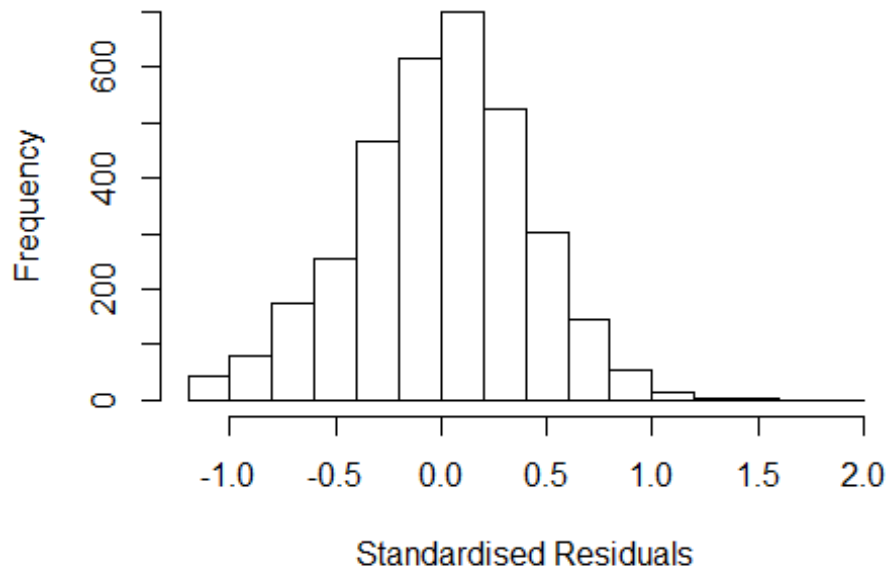
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1092 -0.2711 0.0168 0.2737 1.9790
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05750 0.03750 28.20 <2e-16 ***
## FirstAuthorFemale1 0.03874 0.01526 2.54 0.0112 *
## Year1997 -0.05613 0.05264 -1.07 0.2864
## Year1998 -0.09480 0.04839 -1.96 0.0502 .
## Year1999 -0.12134 0.04944 -2.45 0.0142 *
## Year2000 0.01297 0.05628 0.23 0.8178
## Year2001 -0.05704 0.05760 -0.99 0.3221
## Year2002 -0.02285 0.04527 -0.50 0.6138
## Year2003 -0.11761 0.04685 -2.51 0.0121 *
## Year2004 -0.06526 0.04677 -1.40 0.1630
## Year2005 -0.08557 0.04588 -1.87 0.0623 .
## Year2006 -0.12300 0.04721 -2.61 0.0092 **
```

```

## Year2007          -0.11716      0.04602      -2.55      0.0109 *
## Year2008          -0.08761      0.04598      -1.91      0.0568 .
## Year2009          -0.03514      0.04578      -0.77      0.4428
## Year2010          -0.09855      0.04602      -2.14      0.0323 *
## Year2011          -0.11471      0.04647      -2.47      0.0136 *
## Year2012          -0.00845      0.04645      -0.18      0.8556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0125, Adjusted R-squared:  0.0075
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 3292 is an outlier with |weight| = 0 ( < 3e-05);
## 293 weights are ~= 1. The remaining 3093 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0824 0.8710 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          2.95e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



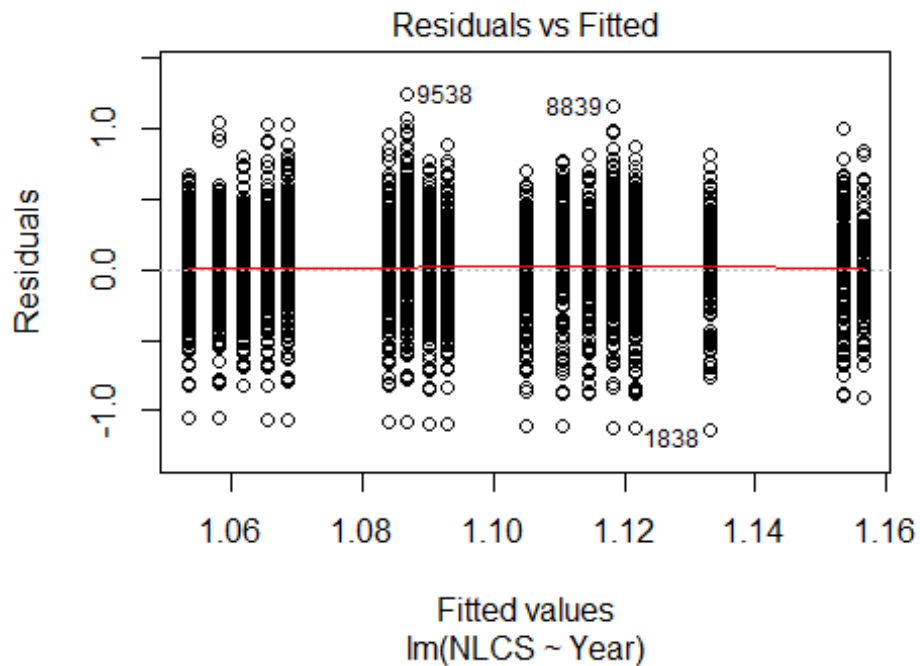
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0856 -0.2671 0.0115 0.2702 1.9607
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.067500 0.037604 28.39 <2e-16 ***
## LastAuthorFemale1 -0.010011 0.017392 -0.58 0.565
## Year1997 -0.054889 0.052796 -1.04 0.299
## Year1998 -0.086901 0.048408 -1.80 0.073 .
## Year1999 -0.120496 0.049482 -2.44 0.015 *
## Year2000 0.018056 0.056891 0.32 0.751
## Year2001 -0.054206 0.057758 -0.94 0.348
## Year2002 -0.018973 0.045285 -0.42 0.675
## Year2003 -0.113013 0.047005 -2.40 0.016 *
## Year2004 -0.060593 0.046885 -1.29 0.196
## Year2005 -0.078209 0.045972 -1.70 0.089 .
## Year2006 -0.111013 0.047207 -2.35 0.019 *
```

```

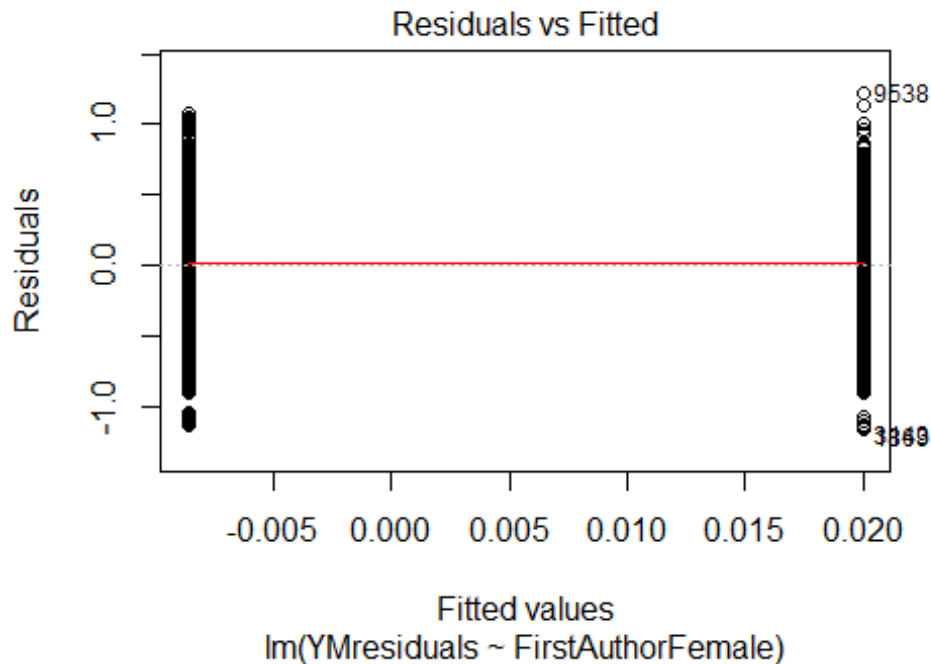
## Year2007          -0.110155    0.046165    -2.39    0.017 *
## Year2008          -0.079489    0.045999    -1.73    0.084 .
## Year2009          -0.025382    0.045729    -0.56    0.579
## Year2010          -0.090052    0.046037    -1.96    0.051 .
## Year2011          -0.107145    0.046533    -2.30    0.021 *
## Year2012          -0.000217    0.046657     0.00    0.996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00552
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 3292 is an outlier with |weight| = 0 ( < 3e-05);
## 278 weights are ~= 1. The remaining 3108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0933 0.8700 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          2.95e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3387"
## [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
## 352 383 385 423 417 418 444 469 455 456 499 527 585 521 558
## 2011 2012
## 555 578
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 154 192 167 177 159 114 268 254 273 282 333 334 357 321 351

```

```
## 2011 2012
## 326 373
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 142 176 152 162 149 105 234 232 251 264 298 294 325 292 318
## 2011 2012
## 295 338
## [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
```

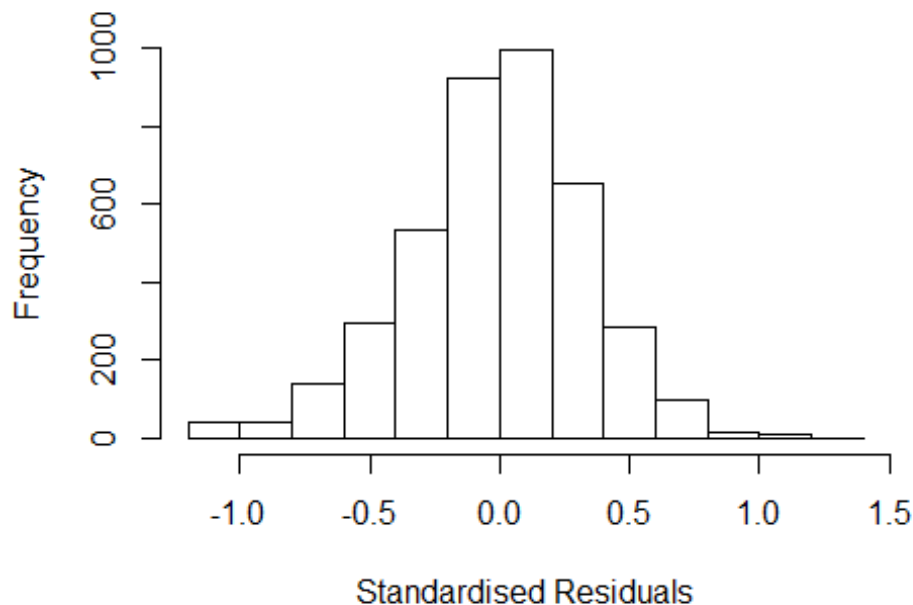


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.8, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.83521363184632"
## [1] "Male first author team size 2018 geometric mean: 3.38925771257944"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.76506124375561"
## [1] "Male last author team size 2018 geometric mean: 3.51169160975218"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.064 1 1.032
## LastAuthorFemale 1.044 1 1.022
## UniqueAuthors 1.158 4 1.018
## Year 1.225 16 1.006
```

Residuals from first and last author and team size



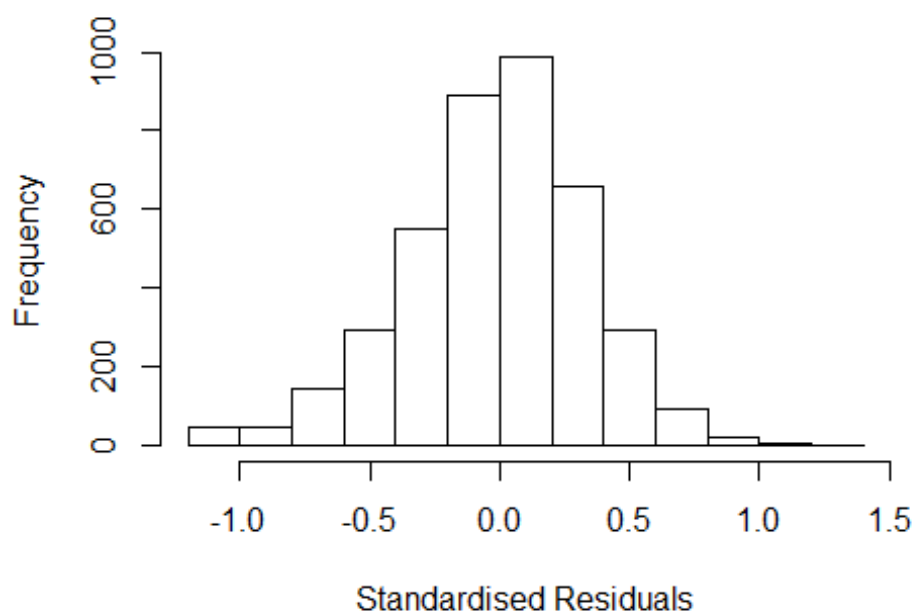
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18076 -0.21345 0.00757 0.21545 1.24102
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09e+00 3.47e-02 31.37 < 2e-16 ***
## FirstAuthorFemale1 1.89e-02 1.17e-02 1.62 0.10582
## LastAuthorFemale1 4.43e-03 1.34e-02 0.33 0.74119
## UniqueAuthors2 6.88e-02 1.91e-02 3.59 0.00033 ***
## UniqueAuthors3 6.72e-02 1.92e-02 3.51 0.00046 ***
## UniqueAuthors4 1.14e-01 2.07e-02 5.49 4.2e-08 ***
## UniqueAuthors5 1.72e-01 2.22e-02 7.73 1.4e-14 ***
## Year1997 -3.32e-02 4.01e-02 -0.83 0.40833
## Year1998 -8.00e-02 4.35e-02 -1.84 0.06587 .
## Year1999 -5.29e-05 4.66e-02 0.00 0.99909
```

```

## Year2000      8.95e-03  4.24e-02  0.21  0.83298
## Year2001      2.27e-02  4.40e-02  0.52  0.60555
## Year2002     -1.19e-02  4.01e-02 -0.30  0.76777
## Year2003     -6.35e-02  3.84e-02 -1.65  0.09828 .
## Year2004     -9.75e-02  3.69e-02 -2.64  0.00832 **
## Year2005     -1.09e-01  3.78e-02 -2.88  0.00394 **
## Year2006     -6.05e-02  3.58e-02 -1.69  0.09141 .
## Year2007     -7.92e-02  3.66e-02 -2.16  0.03050 *
## Year2008     -1.13e-01  3.59e-02 -3.14  0.00167 **
## Year2009     -8.62e-02  3.82e-02 -2.26  0.02411 *
## Year2010     -1.01e-01  3.85e-02 -2.64  0.00840 **
## Year2011     -5.90e-02  3.96e-02 -1.49  0.13642
## Year2012     -9.32e-02  3.87e-02 -2.41  0.01593 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0297, Adjusted R-squared:  0.0243
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 359 weights are ~= 1. The remaining 3668 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0975 0.8610 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      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 1.063 1 1.031
## LastAuthorFemale 1.045 1 1.022
## Year 1.076 16 1.002

```


Residuals from first and last author



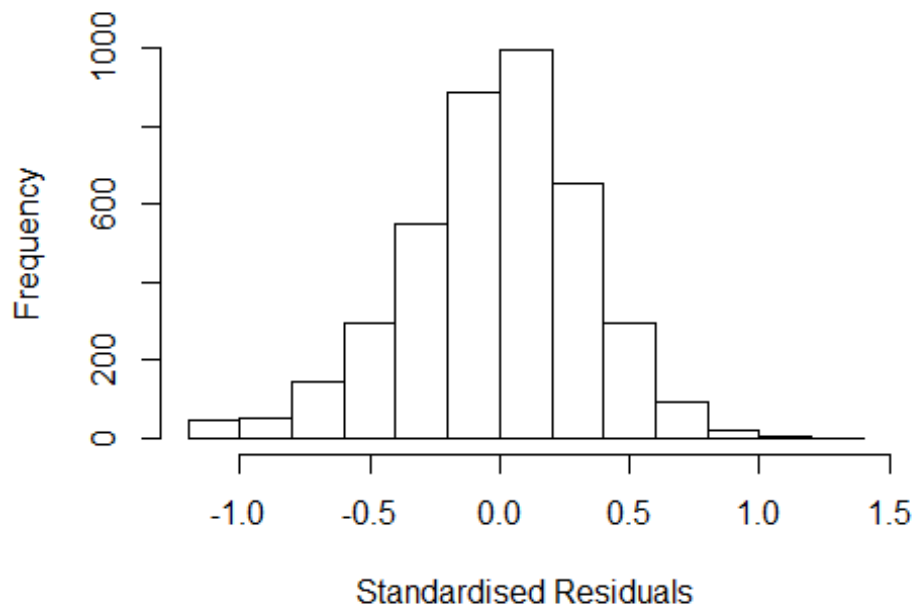
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1853 -0.2159 0.0115 0.2144 1.2069
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14956 0.03224 35.66 <2e-16 ***
## FirstAuthorFemale1 0.02259 0.01175 1.92 0.055 .
## LastAuthorFemale1 0.00363 0.01350 0.27 0.788
## Year1997 -0.03675 0.04007 -0.92 0.359
## Year1998 -0.08671 0.04366 -1.99 0.047 *
## Year1999 -0.00570 0.04685 -0.12 0.903
## Year2000 0.00950 0.04319 0.22 0.826
## Year2001 0.03036 0.04475 0.68 0.497
## Year2002 -0.00817 0.04044 -0.20 0.840
## Year2003 -0.05449 0.03868 -1.41 0.159
## Year2004 -0.08795 0.03736 -2.35 0.019 *
## Year2005 -0.08861 0.03792 -2.34 0.019 *
```

```

## Year2006          -0.04205      0.03608      -1.17      0.244
## Year2007          -0.05420      0.03688      -1.47      0.142
## Year2008          -0.09123      0.03620      -2.52      0.012 *
## Year2009          -0.06433      0.03869      -1.66      0.096 .
## Year2010          -0.07762      0.03870      -2.01      0.045 *
## Year2011          -0.01764      0.03954      -0.45      0.655
## Year2012          -0.05565      0.03909      -1.42      0.155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00634
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 314 weights are ~= 1. The remaining 3713 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.127  0.862  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      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 1.047 1      1.023
## Year      1.047 16      1.001

```

Residuals from first author

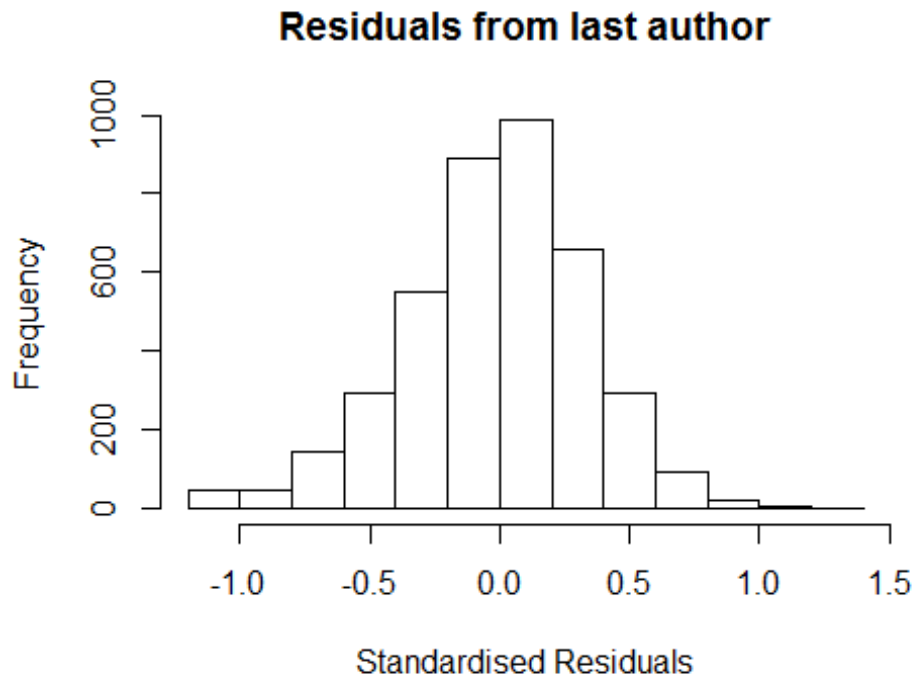


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1827 -0.2159  0.0111  0.2143  1.2093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.15000    0.03219   35.73  <2e-16 ***
## FirstAuthorFemale1 0.02311    0.01167    1.98   0.048 *
## Year1997      -0.03654    0.04004   -0.91   0.362
## Year1998      -0.08656    0.04364   -1.98   0.047 *
## Year1999      -0.00566    0.04684   -0.12   0.904
## Year2000       0.00961    0.04314    0.22   0.824
## Year2001       0.03039    0.04471    0.68   0.497
## Year2002      -0.00797    0.04041   -0.20   0.844
## Year2003      -0.05433    0.03866   -1.41   0.160
## Year2004      -0.08798    0.03735   -2.36   0.019 *
## Year2005      -0.08861    0.03790   -2.34   0.019 *
## Year2006      -0.04159    0.03598   -1.16   0.248
```

```

## Year2007          -0.05403    0.03684   -1.47    0.143
## Year2008          -0.09114    0.03618   -2.52    0.012 *
## Year2009          -0.06422    0.03867   -1.66    0.097 .
## Year2010          -0.07730    0.03862   -2.00    0.045 *
## Year2011          -0.01756    0.03952   -0.44    0.657
## Year2012          -0.05538    0.03904   -1.42    0.156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00656
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 316 weights are ~= 1. The remaining 3711 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.125  0.861  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      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 1.03 1          1.015
## Year              1.03 16          1.001

```



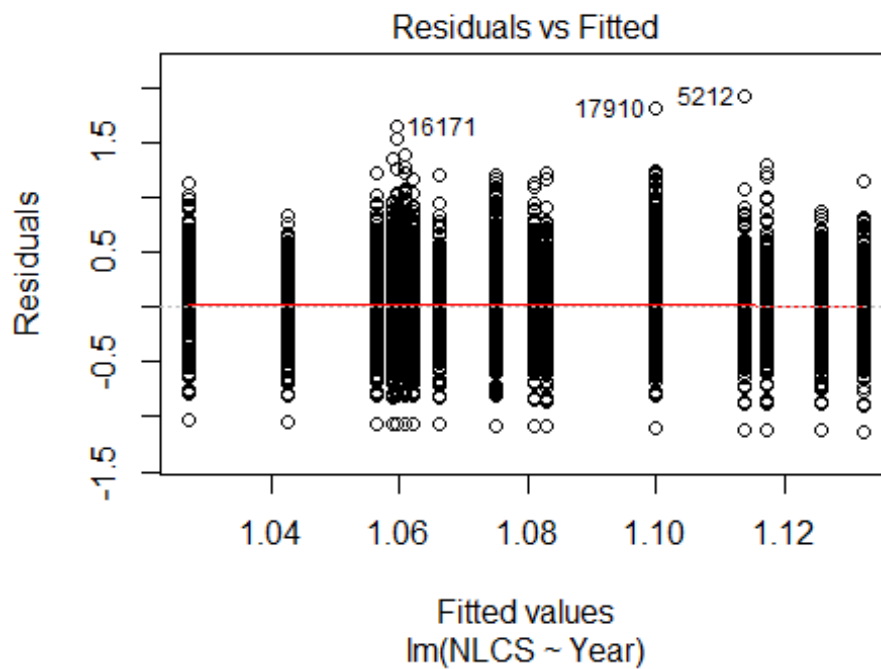
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.171 -0.219 0.010 0.214 1.219
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15315 0.03217 35.84 <2e-16 ***
## LastAuthorFemale1 0.00769 0.01340 0.57 0.566
## Year1997 -0.03695 0.04008 -0.92 0.357
## Year1998 -0.08495 0.04363 -1.95 0.052 .
## Year1999 -0.00268 0.04666 -0.06 0.954
## Year2000 0.00975 0.04312 0.23 0.821
## Year2001 0.03309 0.04470 0.74 0.459
## Year2002 -0.00728 0.04042 -0.18 0.857
## Year2003 -0.05169 0.03860 -1.34 0.181
## Year2004 -0.08691 0.03733 -2.33 0.020 *
## Year2005 -0.08575 0.03781 -2.27 0.023 *
## Year2006 -0.03908 0.03596 -1.09 0.277
```

```

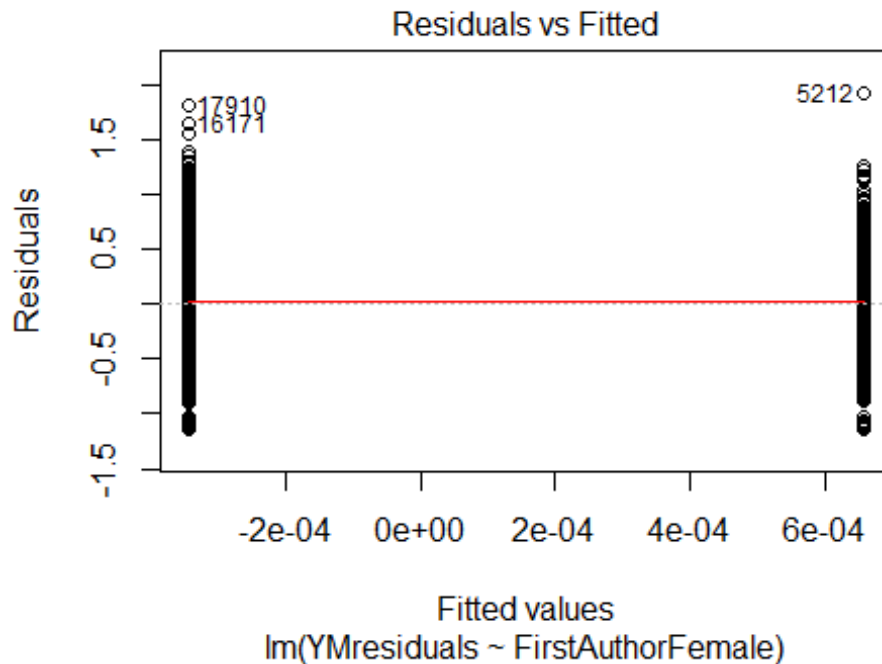
## Year2007          -0.05137      0.03680    -1.40      0.163
## Year2008          -0.08842      0.03614    -2.45      0.014 *
## Year2009          -0.05920      0.03846    -1.54      0.124
## Year2010          -0.07315      0.03860    -1.90      0.058 .
## Year2011          -0.01324      0.03938    -0.34      0.737
## Year2012          -0.05290      0.03902    -1.36      0.175
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.00985,    Adjusted R-squared:  0.00565
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 336 weights are ~= 1. The remaining 3691 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.118  0.862  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      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: 4027"
## [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
##  716  801  754  768  888  866  866  770  769  878  946  961 1042 1228 1218
## 2011 2012
## 1154 1197
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  309  371  373  403  388  327  550  500  494  577  650  698  752  864  880
## 2011 2012

```

```
## 835 901
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 275 344 347 367 348 301 484 460 448 528 597 644 683 781 804
## 2011 2012
## 760 817
## [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
```

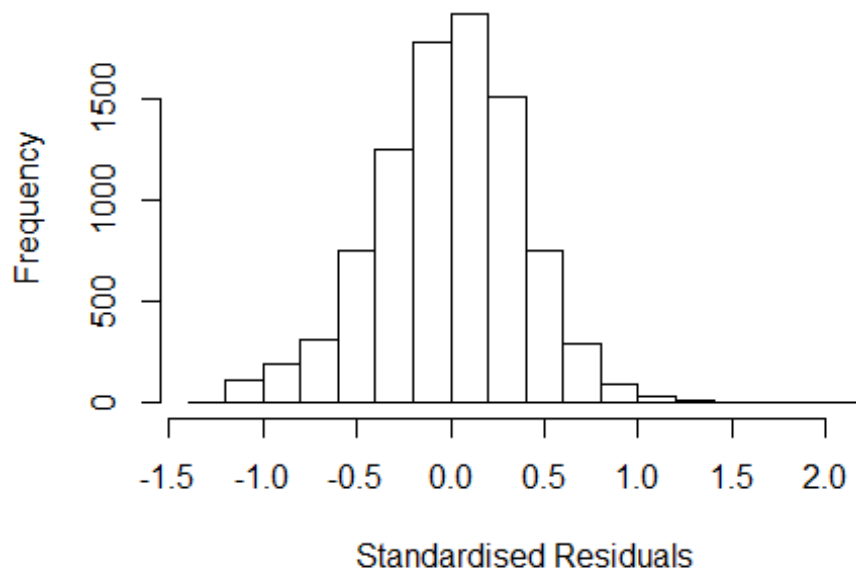


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 17, df = 1, p-value = 4e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.60633383111152"
## [1] "Male first author team size 2018 geometric mean: 3.21147860706392"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 78000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.42998195634946"
## [1] "Male last author team size 2018 geometric mean: 3.35852112762043"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 58000, 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 1.070 1          1.035
## LastAuthorFemale  1.050 1          1.025
## UniqueAuthors     1.138 4          1.016
## Year              1.153 16         1.004
```


Residuals from first and last author and team size



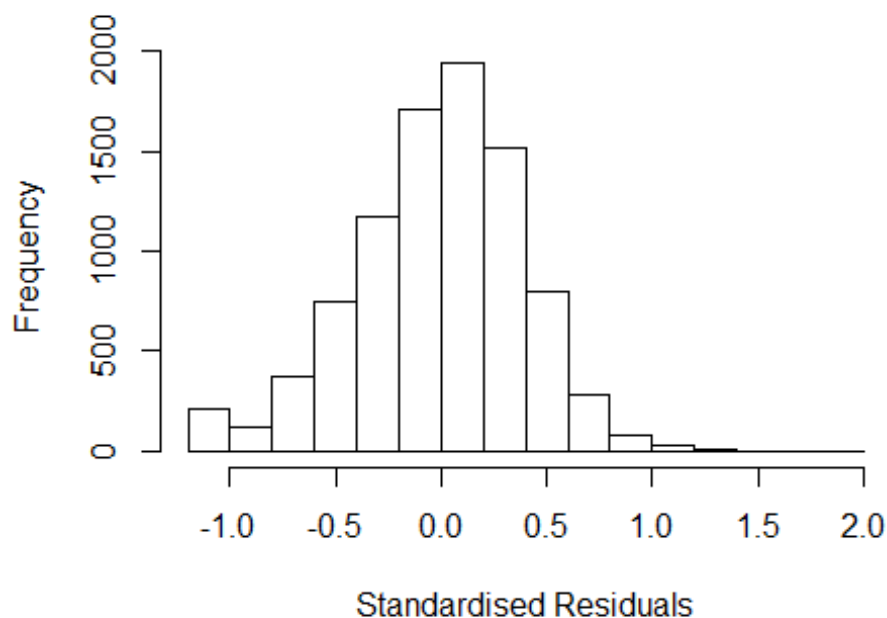
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2440 -0.2487  0.0106  0.2518  2.0881
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01591    0.02283   44.49 < 2e-16 ***
## FirstAuthorFemale1 -0.00987    0.00872   -1.13  0.25803
## LastAuthorFemale1 -0.03127    0.01016   -3.08  0.00210 **
## UniqueAuthors2     0.15519    0.01289   12.04 < 2e-16 ***
## UniqueAuthors3     0.19730    0.01325   14.89 < 2e-16 ***
## UniqueAuthors4     0.25395    0.01493   17.01 < 2e-16 ***
## UniqueAuthors5     0.29689    0.01583   18.75 < 2e-16 ***
## Year1997           0.00416    0.02971    0.14  0.88853
## Year1998          -0.01700    0.03124   -0.54  0.58632
## Year1999          -0.07261    0.02857   -2.54  0.01105 *
```

```

## Year2000      -0.01374      0.02894      -0.47      0.63497
## Year2001      -0.02587      0.03010      -0.86      0.39018
## Year2002      -0.04739      0.02777      -1.71      0.08794 .
## Year2003      -0.09947      0.02769      -3.59      0.00033 ***
## Year2004      -0.10156      0.02691      -3.77      0.00016 ***
## Year2005      -0.10027      0.02670      -3.75      0.00017 ***
## Year2006      -0.13527      0.02635      -5.13      2.9e-07 ***
## Year2007      -0.11696      0.02600      -4.50      6.9e-06 ***
## Year2008      -0.11232      0.02629      -4.27      1.9e-05 ***
## Year2009      -0.11162      0.02621      -4.26      2.1e-05 ***
## Year2010      -0.12046      0.02603      -4.63      3.8e-06 ***
## Year2011      -0.08593      0.02682      -3.20      0.00136 **
## Year2012      -0.10044      0.02759      -3.64      0.00027 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.368
## Multiple R-squared:  0.0623, Adjusted R-squared:  0.06
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 1865 is an outlier with |weight| = 0 ( < 1.1e-05);
## 757 weights are ~= 1. The remaining 8230 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.001  0.869  0.950  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           1.11e-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 1.060 1           1.029
## LastAuthorFemale  1.039 1           1.019
## Year              1.029 16           1.001

```

Residuals from first and last author



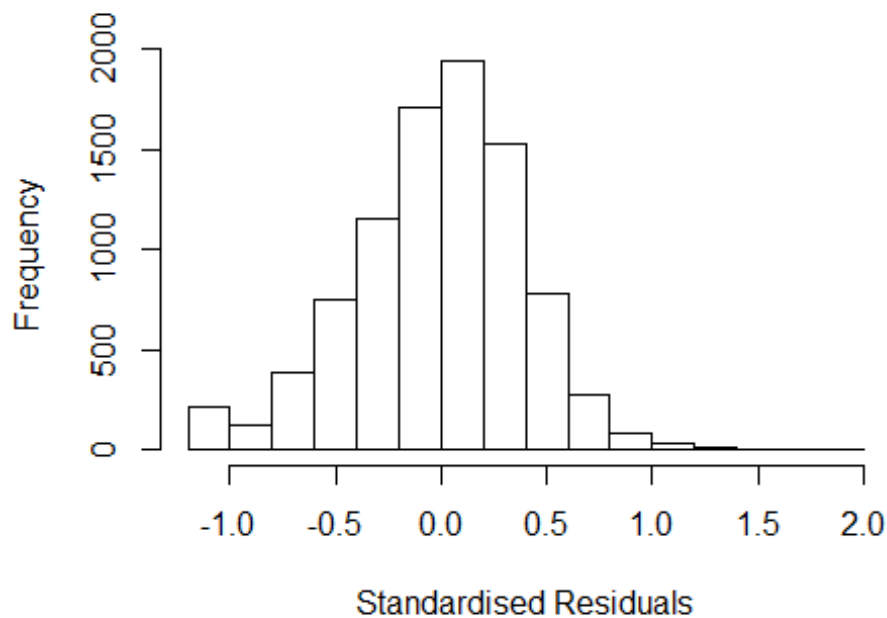
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1458 -0.2564 0.0172 0.2531 1.9453
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13373 0.02163 52.40 < 2e-16 ***
## FirstAuthorFemale1 0.00599 0.00885 0.68 0.4986
## LastAuthorFemale1 -0.04468 0.01034 -4.32 1.6e-05 ***
## Year1997 0.00810 0.02956 0.27 0.7840
## Year1998 -0.02489 0.03087 -0.81 0.4201
## Year1999 -0.07820 0.02903 -2.69 0.0071 **
## Year2000 0.00608 0.02902 0.21 0.8342
## Year2001 -0.00329 0.02976 -0.11 0.9119
## Year2002 -0.02227 0.02742 -0.81 0.4168
## Year2003 -0.06181 0.02797 -2.21 0.0271 *
## Year2004 -0.06270 0.02710 -2.31 0.0207 *
## Year2005 -0.05998 0.02690 -2.23 0.0258 *
```

```

## Year2006      -0.08672    0.02664   -3.26   0.0011 **
## Year2007      -0.05618    0.02615   -2.15   0.0317 *
## Year2008      -0.05182    0.02625   -1.97   0.0484 *
## Year2009      -0.05475    0.02632   -2.08   0.0376 *
## Year2010      -0.05480    0.02614   -2.10   0.0361 *
## Year2011      -0.02159    0.02690   -0.80   0.4222
## Year2012      -0.02291    0.02762   -0.83   0.4069
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.00692,    Adjusted R-squared:  0.00493
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1865 is an outlier with |weight| = 0 ( < 1.1e-05);
## 756 weights are ~= 1. The remaining 8231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8680 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      1.11e-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 1.023 1          1.011
## Year              1.023 16          1.001

```

Residuals from first author



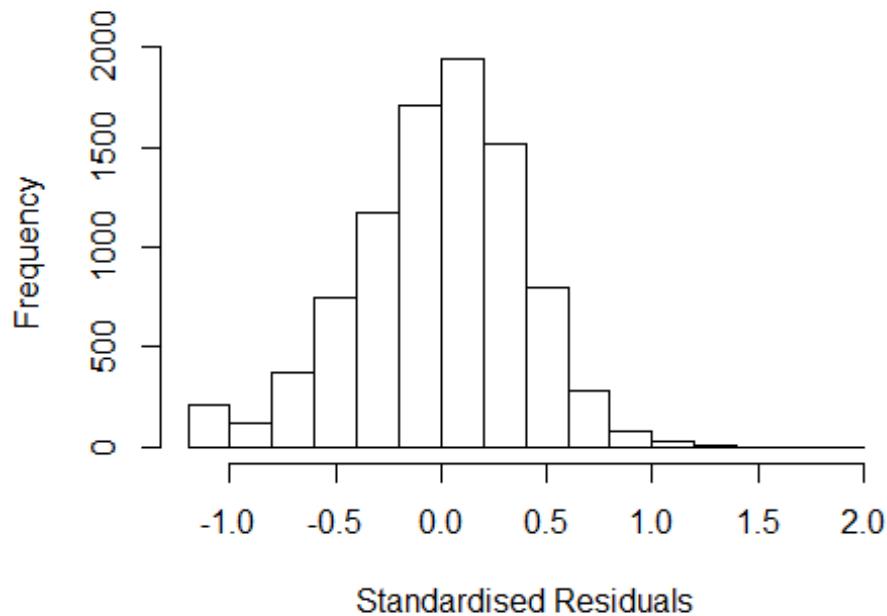
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1333 -0.2570 0.0158 0.2539 1.9153
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12739 0.02155 52.30 < 2e-16 ***
## FirstAuthorFemale1 -0.00288 0.00874 -0.33 0.74174
## Year1997 0.00595 0.02956 0.20 0.84040
## Year1998 -0.02603 0.03089 -0.84 0.39948
## Year1999 -0.07766 0.02904 -2.67 0.00750 **
## Year2000 0.00549 0.02901 0.19 0.84994
## Year2001 -0.00279 0.02981 -0.09 0.92532
## Year2002 -0.02316 0.02741 -0.84 0.39821
## Year2003 -0.06202 0.02791 -2.22 0.02631 *
## Year2004 -0.06274 0.02706 -2.32 0.02045 *
## Year2005 -0.06054 0.02688 -2.25 0.02433 *
## Year2006 -0.08835 0.02663 -3.32 0.00091 ***
```

```

## Year2007          -0.05695      0.02616      -2.18      0.02950 *
## Year2008          -0.05131      0.02622      -1.96      0.05042 .
## Year2009          -0.05597      0.02628      -2.13      0.03323 *
## Year2010          -0.05605      0.02610      -2.15      0.03174 *
## Year2011          -0.02160      0.02690      -0.80      0.42199
## Year2012          -0.02317      0.02757      -0.84      0.40087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.0047, Adjusted R-squared:  0.00281
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1865 is an outlier with |weight| = 0 ( < 1.1e-05);
## 778 weights are ~= 1. The remaining 8209 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0146 0.8670 0.9490 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          1.11e-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 1.004 1          1.002
## Year          1.004 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
## -1.1429 -0.2565 0.0168 0.2532 1.9482
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13469 0.02159 52.56 < 2e-16 ***
## LastAuthorFemale1 -0.04313 0.01016 -4.24 2.2e-05 ***
## Year1997 0.00818 0.02955 0.28 0.7819
## Year1998 -0.02433 0.03081 -0.79 0.4299
## Year1999 -0.07783 0.02901 -2.68 0.0073 **
## Year2000 0.00638 0.02902 0.22 0.8259
## Year2001 -0.00277 0.02973 -0.09 0.9259
## Year2002 -0.02172 0.02739 -0.79 0.4277
## Year2003 -0.06108 0.02790 -2.19 0.0286 *
## Year2004 -0.06196 0.02705 -2.29 0.0220 *
## Year2005 -0.05924 0.02684 -2.21 0.0274 *
## Year2006 -0.08585 0.02657 -3.23 0.0012 **
```

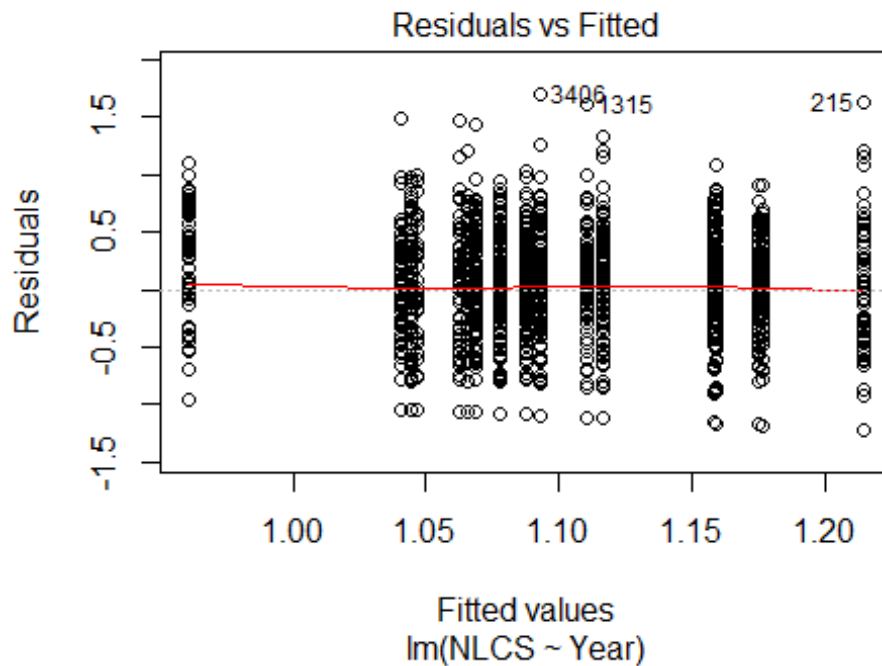
```

## Year2007          -0.05529      0.02610      -2.12      0.0341 *
## Year2008          -0.05071      0.02614      -1.94      0.0525 .
## Year2009          -0.05371      0.02624      -2.05      0.0407 *
## Year2010          -0.05378      0.02605      -2.06      0.0390 *
## Year2011          -0.02038      0.02677      -0.76      0.4465
## Year2012          -0.02183      0.02752      -0.79      0.4275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.00687,    Adjusted R-squared:  0.00499
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1865 is an outlier with |weight| = 0 ( < 1.1e-05);
## 756 weights are ~= 1. The remaining 8231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0069 0.8680 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          1.11e-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: 8988"
## [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
## 151 147 144 151 319 170 185 185 166 141 181 231 277 318 342
## 2011 2012
## 318 310
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 56 45 62 71 59 90 75 81 78 77 97 121 110 132

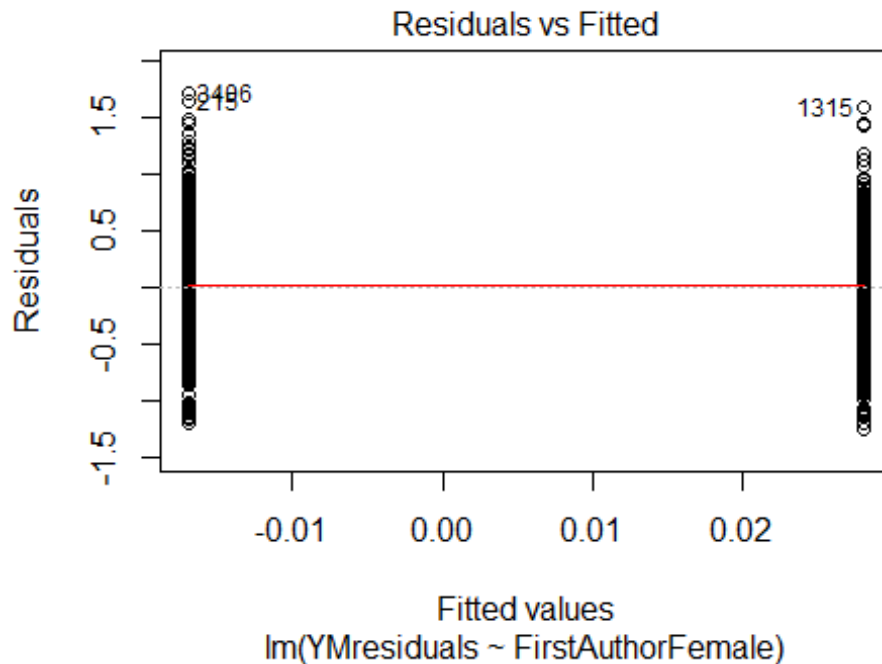
```



```
## 2011 2012
## 144 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 48 41 56 63 50 82 65 72 71 67 86 111 96 118
## 2011 2012
## 129 118
## [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
```

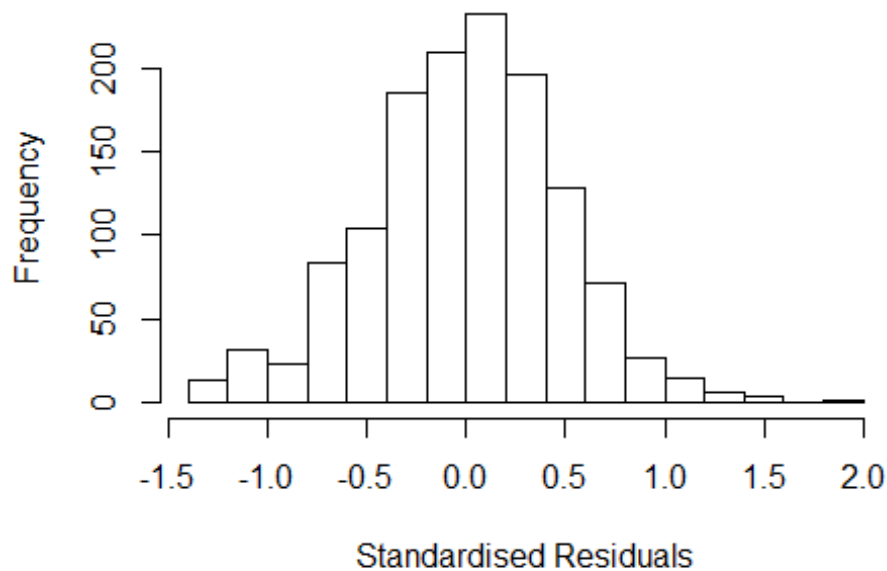


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.33, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 4.72443786045662"
## [1] "Male first author team size 2018 geometric mean: 3.79143078444924"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7600, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.50925057391744"
## [1] "Male last author team size 2018 geometric mean: 4.24408805021401"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6800, 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 1.101 1      1.049
## LastAuthorFemale  1.077 1      1.038
## UniqueAuthors    1.286 4      1.032
## Year              1.387 16     1.010
```

Residuals from first and last author and team size



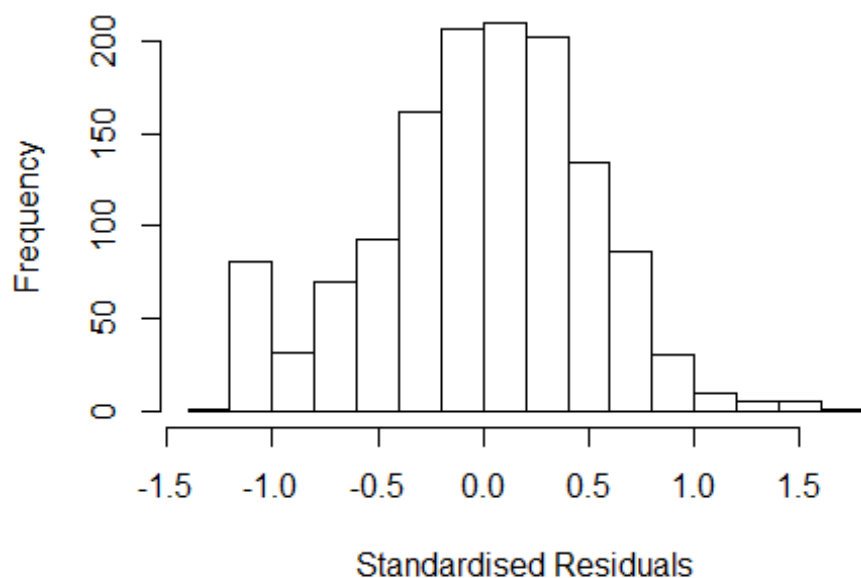
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3361 -0.3103  0.0109  0.3082  1.9218
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.71402    0.08390   8.51  < 2e-16 ***
## FirstAuthorFemale1 0.01006    0.02798   0.36   0.719
## LastAuthorFemale1 -0.00998    0.03199  -0.31   0.755
## UniqueAuthors2    0.39541    0.05168   7.65 3.8e-14 ***
## UniqueAuthors3    0.45621    0.04934   9.25 < 2e-16 ***
## UniqueAuthors4    0.59761    0.05018  11.91 < 2e-16 ***
## UniqueAuthors5    0.60040    0.04986  12.04 < 2e-16 ***
## Year1997          0.05099    0.11558   0.44   0.659
## Year1998         -0.03713    0.11386  -0.33   0.744
## Year1999         -0.07460    0.09914  -0.75   0.452
```

```

## Year2000      -0.13151    0.10459   -1.26    0.209
## Year2001      0.01438    0.10920    0.13    0.895
## Year2002     -0.02895    0.08520   -0.34    0.734
## Year2003     -0.10986    0.09250   -1.19    0.235
## Year2004     -0.03495    0.08707   -0.40    0.688
## Year2005     -0.02740    0.08636   -0.32    0.751
## Year2006     -0.17433    0.09649   -1.81    0.071 .
## Year2007     -0.01075    0.09120   -0.12    0.906
## Year2008     -0.07169    0.08668   -0.83    0.408
## Year2009     -0.01639    0.08939   -0.18    0.855
## Year2010     -0.06427    0.08525   -0.75    0.451
## Year2011     -0.06706    0.08420   -0.80    0.426
## Year2012     -0.05233    0.08576   -0.61    0.542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.167, Adjusted R-squared:  0.153
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1210 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0389 0.8600 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          7.52e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.052 1 1.026
## LastAuthorFemale 1.055 1 1.027
## Year 1.102 16 1.003

```

Residuals from first and last author



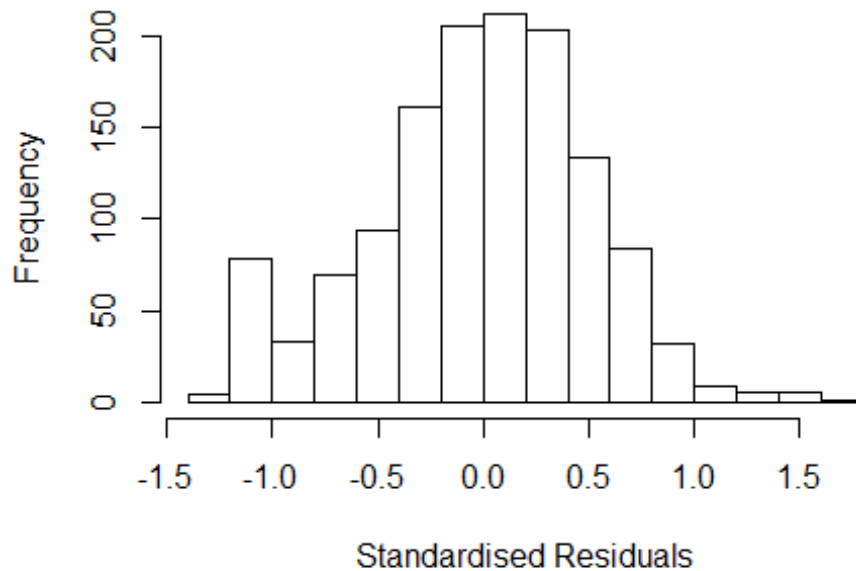
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2087 -0.3219  0.0225  0.3331  1.7071
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14409    0.07344   15.58  <2e-16 ***
## FirstAuthorFemale1  0.05669    0.03003    1.89   0.059 .
## LastAuthorFemale1 -0.01646    0.03563   -0.46   0.644
## Year1997        -0.01217    0.12042   -0.10   0.920
## Year1998        -0.05583    0.12337   -0.45   0.651
## Year1999        -0.10397    0.09919   -1.05   0.295
## Year2000        -0.16230    0.12460   -1.30   0.193
## Year2001        -0.10136    0.11904   -0.85   0.395
## Year2002        -0.03116    0.08940   -0.35   0.728
## Year2003        -0.12476    0.09613   -1.30   0.195
## Year2004         0.00960    0.08791    0.11   0.913
## Year2005         0.01747    0.08947    0.20   0.845
```

```

## Year2006      -0.14967    0.09581   -1.56    0.118
## Year2007      0.00789    0.09212    0.09    0.932
## Year2008     -0.09493    0.08766   -1.08    0.279
## Year2009     -0.02454    0.09347   -0.26    0.793
## Year2010     -0.09509    0.08656   -1.10    0.272
## Year2011     -0.08530    0.08687   -0.98    0.326
## Year2012     -0.04125    0.08614   -0.48    0.632
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.0136, Adjusted R-squared:  2.52e-05
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 103 weights are ~= 1. The remaining 1227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.859  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      7.52e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.05 1      1.025
## Year      1.05 16      1.002

```

Residuals from first author



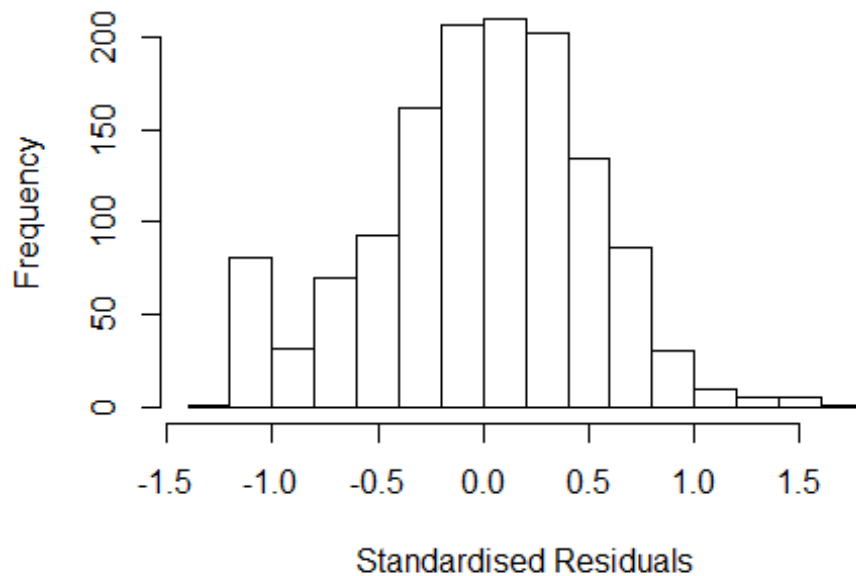
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2039 -0.3243 0.0191 0.3339 1.7086
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14088 0.07290 15.65 <2e-16 ***
## FirstAuthorFemale1 0.05401 0.03031 1.78 0.075 .
## Year1997 -0.01051 0.12039 -0.09 0.930
## Year1998 -0.05716 0.12328 -0.46 0.643
## Year1999 -0.10223 0.09929 -1.03 0.303
## Year2000 -0.16137 0.12455 -1.30 0.195
## Year2001 -0.10179 0.11912 -0.85 0.393
## Year2002 -0.03147 0.08966 -0.35 0.726
## Year2003 -0.12509 0.09623 -1.30 0.194
## Year2004 0.01027 0.08797 0.12 0.907
## Year2005 0.01920 0.08942 0.21 0.830
## Year2006 -0.14854 0.09572 -1.55 0.121
```

```

## Year2007          0.00899    0.09201    0.10    0.922
## Year2008          -0.09446    0.08782   -1.08    0.282
## Year2009          -0.02345    0.09352   -0.25    0.802
## Year2010          -0.09454    0.08666   -1.09    0.276
## Year2011          -0.08583    0.08716   -0.98    0.325
## Year2012          -0.04082    0.08626   -0.47    0.636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.000675
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1226 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.860  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      7.52e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



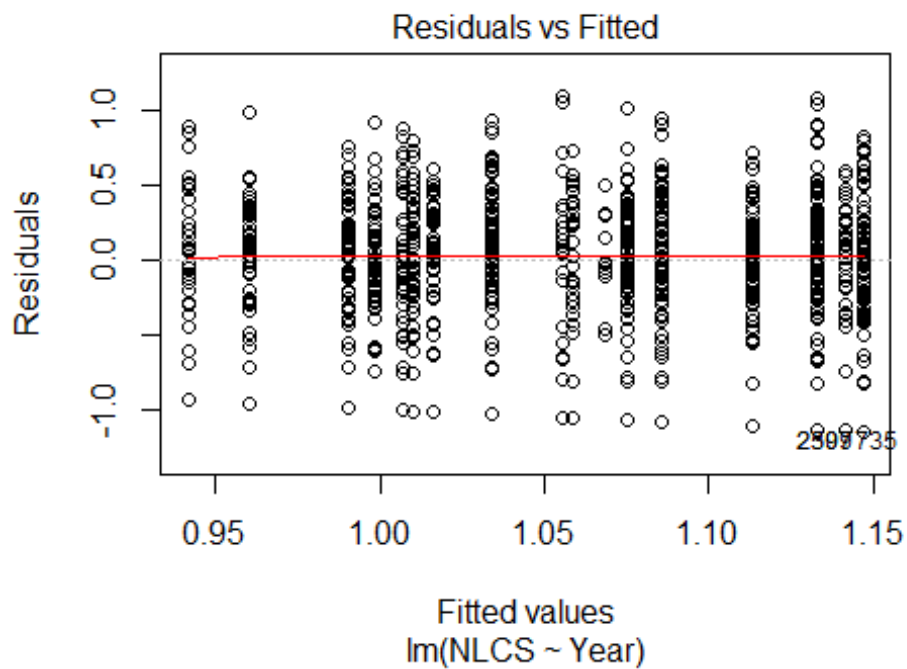
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1822 -0.3151 0.0251 0.3324 1.6955
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15088 0.07314 15.73 <2e-16 ***
## LastAuthorFemale1 -0.00425 0.03553 -0.12 0.90
## Year1997 -0.00736 0.12001 -0.06 0.95
## Year1998 -0.05376 0.12419 -0.43 0.67
## Year1999 -0.09967 0.09906 -1.01 0.31
## Year2000 -0.15502 0.12518 -1.24 0.22
## Year2001 -0.09436 0.11727 -0.80 0.42
## Year2002 -0.01774 0.08879 -0.20 0.84
## Year2003 -0.10794 0.09594 -1.13 0.26
## Year2004 0.02368 0.08773 0.27 0.79
## Year2005 0.03132 0.08879 0.35 0.72
## Year2006 -0.13795 0.09523 -1.45 0.15
```

```

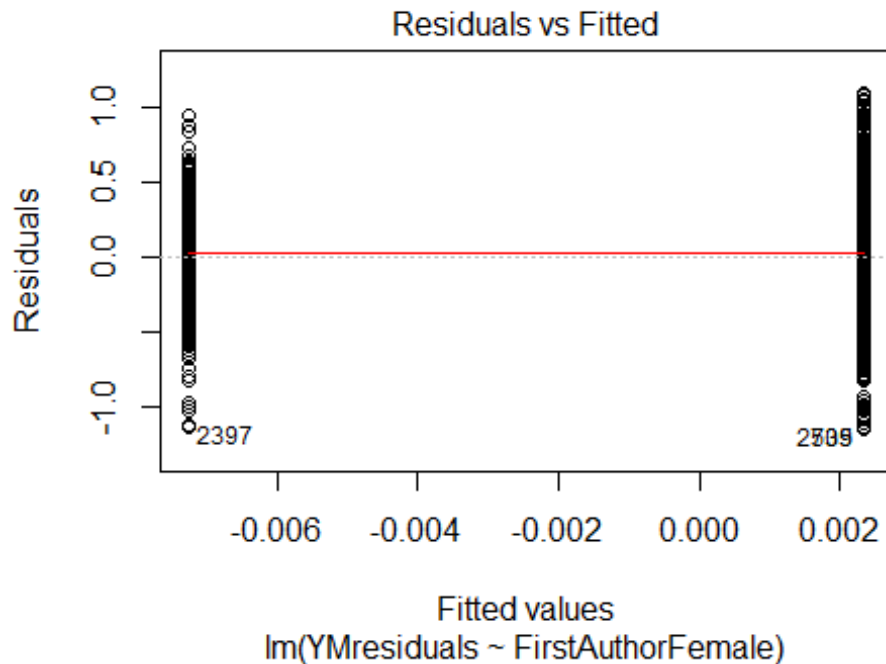
## Year2007          0.02372      0.09136      0.26      0.80
## Year2008          -0.07954      0.08739     -0.91      0.36
## Year2009          -0.01196      0.09307     -0.13      0.90
## Year2010          -0.08450      0.08641     -0.98      0.33
## Year2011          -0.07248      0.08671     -0.84      0.40
## Year2012          -0.02506      0.08554     -0.29      0.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.011, Adjusted R-squared:  -0.00182
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.196  0.859  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      7.52e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1330"
## [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
##   98   92  122  117   91  119  113  123  111  100  146  107  144  140  152
## 2011 2012
##  155  156
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   11   37   47   29   46   42   61   57   57   89   69   89   88   96
## 2011 2012

```

```
## 103 94
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 11 34 45 27 43 38 56 47 50 80 55 74 81 83
## 2011 2012
## 98 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 = 33, df = 16, p-value = 0.007
```

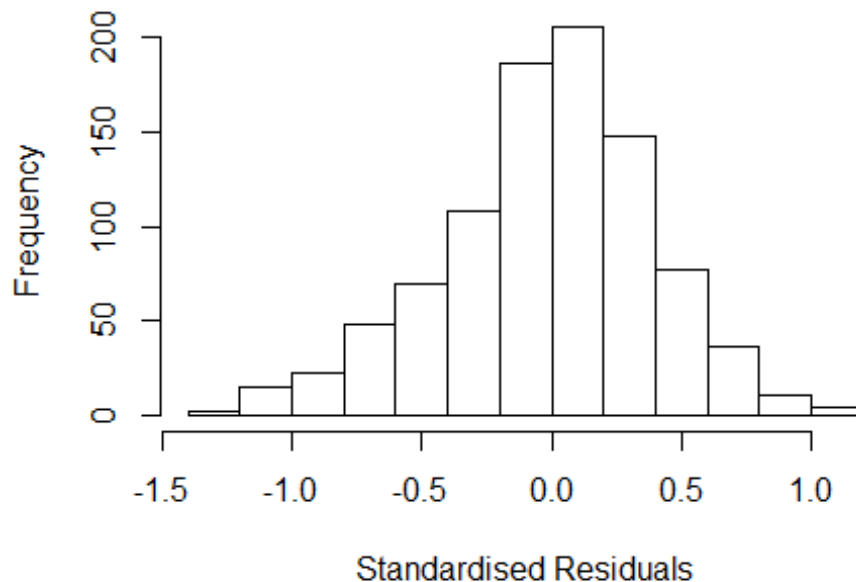


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.52146486114552"
## [1] "Male first author team size 2018 geometric mean: 3.41188248847895"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 900, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.79077082422627"
## [1] "Male last author team size 2018 geometric mean: 3.3705937620959"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 720, 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 1.258 1      1.122
## LastAuthorFemale  1.220 1      1.105
## UniqueAuthors     1.525 4      1.054
## Year              1.753 16      1.018
```

Residuals from first and last author and team size



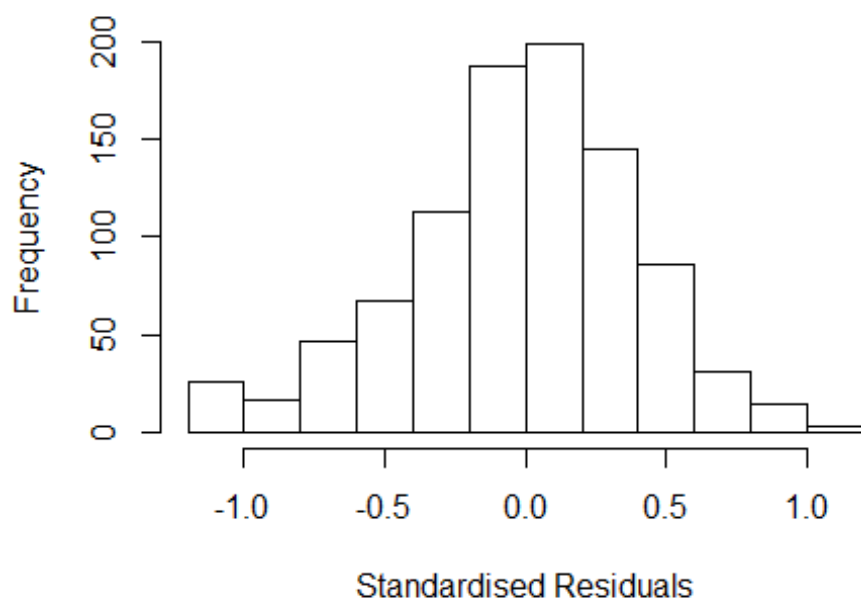
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2926 -0.2484 0.0159 0.2486 1.1730
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9760 0.1357 7.19 1.3e-12 ***
## FirstAuthorFemale1 -0.0257 0.0332 -0.77 0.43901
## LastAuthorFemale1 -0.0137 0.0378 -0.36 0.71751
## UniqueAuthors2 0.0569 0.0445 1.28 0.20122
## UniqueAuthors3 0.1604 0.0439 3.65 0.00027 ***
## UniqueAuthors4 0.2233 0.0491 4.54 6.2e-06 ***
## UniqueAuthors5 0.2059 0.0484 4.25 2.3e-05 ***
## Year1997 0.0434 0.1606 0.27 0.78697
## Year1998 -0.0654 0.1555 -0.42 0.67394
## Year1999 -0.0593 0.1585 -0.37 0.70836
```

```

## Year2000          0.0250      0.1559      0.16  0.87260
## Year2001          0.0992      0.1430      0.69  0.48817
## Year2002         -0.0272      0.1478     -0.18  0.85388
## Year2003         -0.1244      0.1433     -0.87  0.38553
## Year2004         -0.0474      0.1489     -0.32  0.75056
## Year2005         -0.0781      0.1430     -0.55  0.58526
## Year2006          0.0298      0.1421      0.21  0.83415
## Year2007         -0.0822      0.1468     -0.56  0.57536
## Year2008          0.0577      0.1391      0.41  0.67838
## Year2009          0.0933      0.1446      0.65  0.51891
## Year2010         -0.0217      0.1432     -0.15  0.87969
## Year2011         -0.0798      0.1401     -0.57  0.56892
## Year2012          0.0471      0.1413      0.33  0.73870
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.372
## Multiple R-squared:  0.071, Adjusted R-squared:  0.0486
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 849 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.204  0.856  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      1.07e-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 1.226 1      1.107
## LastAuthorFemale  1.196 1      1.094
## Year              1.182 16      1.005

```

Residuals from first and last author



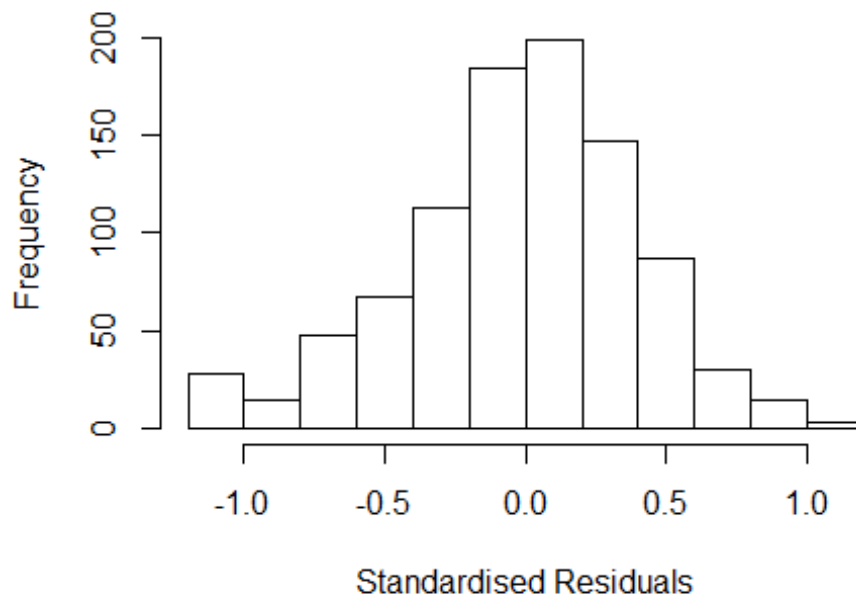
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1942 -0.2596  0.0103  0.2482  1.0964
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05259    0.12945   8.13 1.4e-15 ***
## FirstAuthorFemale1 -0.01217    0.03358  -0.36  0.72
## LastAuthorFemale1 -0.02442    0.03814  -0.64  0.52
## Year1997         0.02312    0.15812   0.15  0.88
## Year1998        -0.08504    0.15386  -0.55  0.58
## Year1999        -0.03761    0.15008  -0.25  0.80
## Year2000         0.07394    0.15171   0.49  0.63
## Year2001         0.10128    0.13649   0.74  0.46
## Year2002        -0.00358    0.14192  -0.03  0.98
## Year2003        -0.09714    0.13873  -0.70  0.48
## Year2004        -0.01468    0.14358  -0.10  0.92
## Year2005        -0.05935    0.13842  -0.43  0.67
```

```

## Year2006          0.06932    0.13687    0.51    0.61
## Year2007          -0.02570    0.14085   -0.18    0.86
## Year2008          0.09285    0.13416    0.69    0.49
## Year2009          0.14160    0.13996    1.01    0.31
## Year2010          0.03943    0.13753    0.29    0.77
## Year2011          -0.01506    0.13415   -0.11    0.91
## Year2012          0.11915    0.13543    0.88    0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0319, Adjusted R-squared:  0.0128
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.305  0.862  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      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 1.092 1      1.045
## Year              1.092 16      1.003

```


Residuals from first author



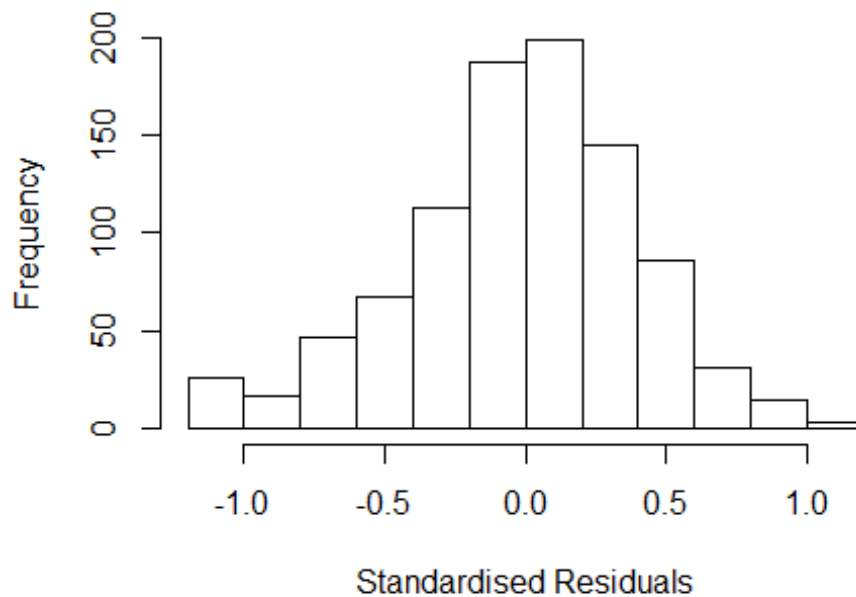
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1927 -0.2571 0.0096 0.2483 1.0996
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04942 0.12922 8.12 1.5e-15 ***
## FirstAuthorFemale1 -0.01740 0.03189 -0.55 0.59
## Year1997 0.02390 0.15798 0.15 0.88
## Year1998 -0.08406 0.15392 -0.55 0.59
## Year1999 -0.03725 0.14975 -0.25 0.80
## Year2000 0.07440 0.15206 0.49 0.62
## Year2001 0.10226 0.13639 0.75 0.45
## Year2002 -0.00193 0.14184 -0.01 0.99
## Year2003 -0.09381 0.13844 -0.68 0.50
## Year2004 -0.01405 0.14350 -0.10 0.92
## Year2005 -0.05970 0.13845 -0.43 0.67
## Year2006 0.06983 0.13688 0.51 0.61
```

```

## Year2007      -0.02599    0.14088   -0.18    0.85
## Year2008      0.09195    0.13410    0.69    0.49
## Year2009      0.14327    0.13981    1.02    0.31
## Year2010      0.04064    0.13747    0.30    0.77
## Year2011     -0.01569    0.13416   -0.12    0.91
## Year2012      0.11836    0.13544    0.87    0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0136
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 850 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.305  0.862  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      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 1.067 1          1.033
## Year            1.067 16          1.002

```

Residuals from last author



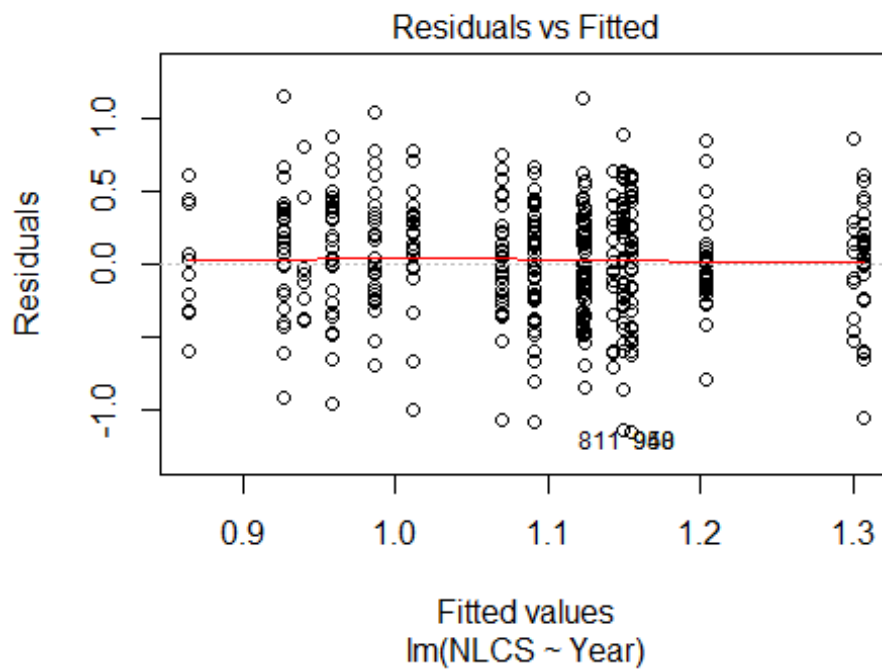
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19027 -0.26006  0.00468  0.25148  1.09893
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05007    0.12879   8.15 1.2e-15 ***
## LastAuthorFemale1 -0.02790    0.03615  -0.77   0.44
## Year1997         0.02476    0.15794   0.16   0.88
## Year1998        -0.08371    0.15361  -0.54   0.59
## Year1999        -0.03602    0.14978  -0.24   0.81
## Year2000         0.07301    0.15171   0.48   0.63
## Year2001         0.10220    0.13629   0.75   0.45
## Year2002        -0.00323    0.14182  -0.02   0.98
## Year2003        -0.09741    0.13859  -0.70   0.48
## Year2004        -0.01366    0.14334  -0.10   0.92
## Year2005        -0.05853    0.13814  -0.42   0.67
## Year2006         0.06841    0.13679   0.50   0.62
```

```

## Year2007          -0.02509      0.14065      -0.18      0.86
## Year2008           0.09299      0.13404       0.69      0.49
## Year2009           0.14019      0.13993       1.00      0.32
## Year2010           0.03891      0.13750       0.28      0.78
## Year2011          -0.01485      0.13395      -0.11      0.91
## Year2012           0.12006      0.13519       0.89      0.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0317, Adjusted R-squared:  0.0137
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.863  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      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: 934"
## [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
##   65   60   71   63   68   64   83   77   77   81   90   79   99   84   79
## 2011 2012
##  103   83
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10    8   12   20   27   18   31   27   37   34   38   41   36   35   31
## 2011 2012

```

```
## 33 33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 7 10 18 24 17 23 25 32 32 34 35 32 35 28
## 2011 2012
## 31 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 = 19, df = 16, p-value = 0.3
```



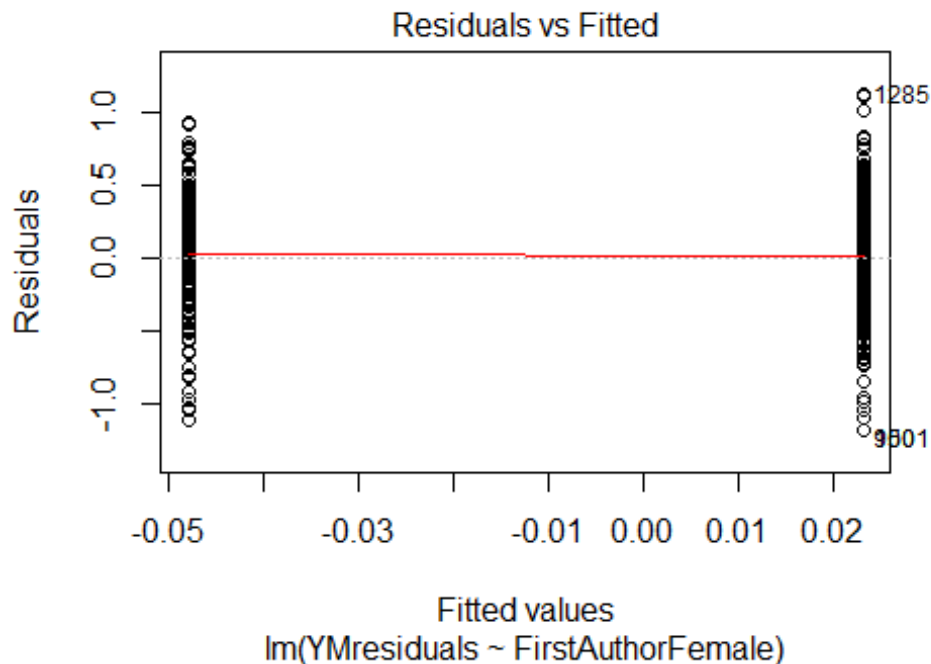
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.89, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 3.7224194364084"
## [1] "Male first author team size 2018 geometric mean: 3.58957314694989"

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

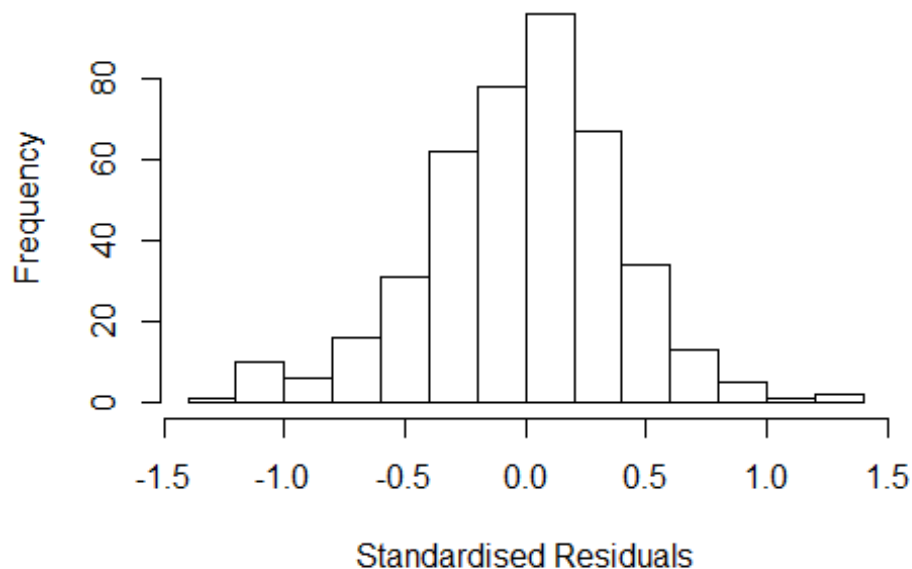
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 94, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.47807902340483"
## [1] "Male last author team size 2018 geometric mean: 3.43788546975182"

## 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 = 0.2
## 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 1.234 1      1.111
## LastAuthorFemale  1.262 1      1.123
## UniqueAuthors    2.196 4      1.103
## Year              2.589 16     1.030
```

Residuals from first and last author and team size



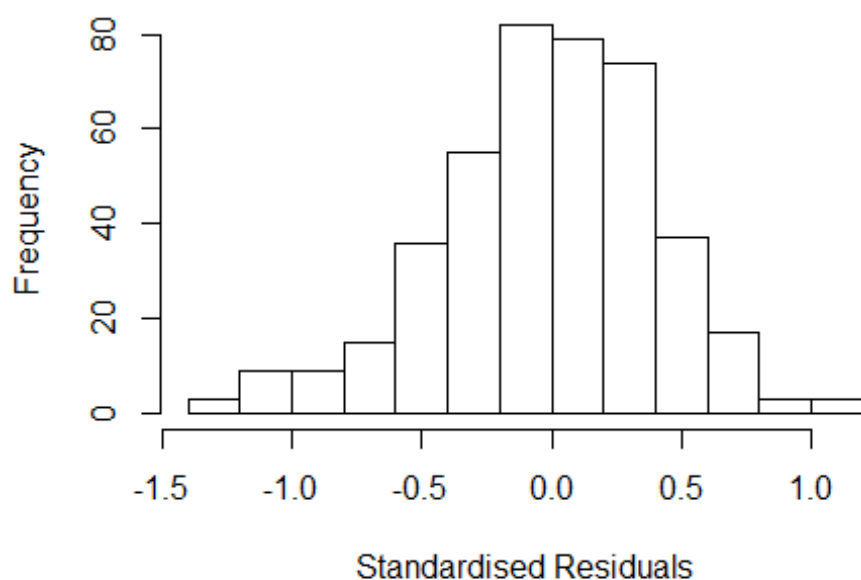
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.209 -0.256 0.011 0.243 1.218
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.69549 0.16279 4.27 2.4e-05 ***
## FirstAuthorFemale1 -0.08421 0.04473 -1.88 0.06046 .
## LastAuthorFemale1 0.00427 0.04715 0.09 0.92795
## UniqueAuthors2 0.16973 0.08916 1.90 0.05766 .
## UniqueAuthors3 0.09408 0.08953 1.05 0.29398
## UniqueAuthors4 0.22222 0.08730 2.55 0.01129 *
## UniqueAuthors5 0.29987 0.08685 3.45 0.00061 ***
## Year1997 0.08332 0.21037 0.40 0.69225
## Year1998 0.49471 0.18516 2.67 0.00785 **
## Year1999 0.31764 0.17902 1.77 0.07677 .
```

```

## Year2000          0.48672    0.15370    3.17  0.00166 **
## Year2001          0.29369    0.17091    1.72  0.08649 .
## Year2002          0.31124    0.15993    1.95  0.05234 .
## Year2003          0.37035    0.15458    2.40  0.01704 *
## Year2004          0.25319    0.15195    1.67  0.09643 .
## Year2005          0.36448    0.16257    2.24  0.02551 *
## Year2006          0.28635    0.15124    1.89  0.05904 .
## Year2007          0.39476    0.16903    2.34  0.02002 *
## Year2008          0.12956    0.16286    0.80  0.42675
## Year2009          0.16360    0.16374    1.00  0.31833
## Year2010          0.06394    0.17082    0.37  0.70836
## Year2011          0.29176    0.16299    1.79  0.07421 .
## Year2012          0.21290    0.15535    1.37  0.17131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.131, Adjusted R-squared:  0.0826
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 386 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.872  0.949   0.891   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      2.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.154 1      1.074
## LastAuthorFemale  1.184 1      1.088
## Year              1.362 16      1.010

```


Residuals from first and last author



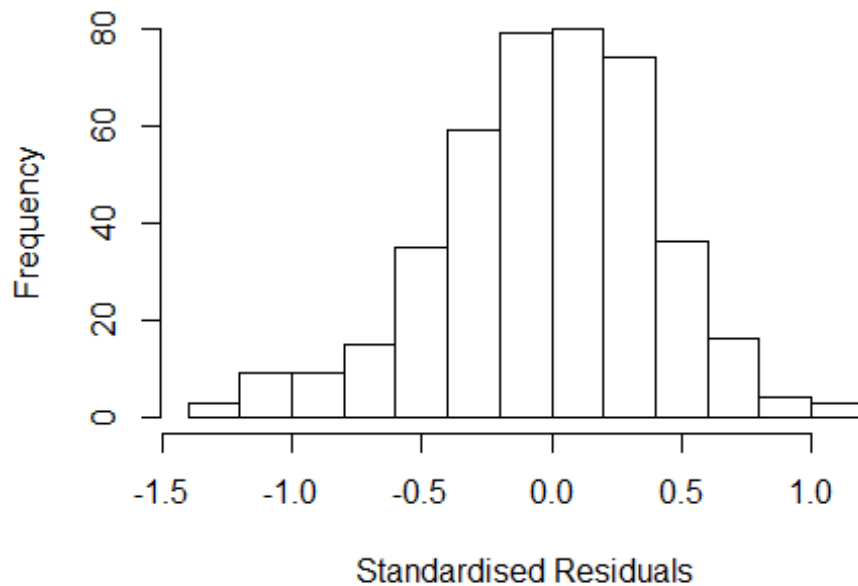
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26504 -0.25845  0.00541  0.25355  1.12888
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8672    0.1376   6.30 7.6e-10 ***
## FirstAuthorFemale1 -0.0596    0.0454  -1.31  0.1905
## LastAuthorFemale1 -0.0135    0.0463  -0.29  0.7709
## Year1997         0.1128    0.2062   0.55  0.5846
## Year1998         0.4931    0.1774   2.78  0.0057 **
## Year1999         0.2739    0.1664   1.65  0.1005
## Year2000         0.4708    0.1495   3.15  0.0018 **
## Year2001         0.2973    0.1718   1.73  0.0842 .
## Year2002         0.2768    0.1561   1.77  0.0769 .
## Year2003         0.3601    0.1498   2.40  0.0167 *
## Year2004         0.2671    0.1490   1.79  0.0738 .
## Year2005         0.3528    0.1602   2.20  0.0282 *
```

```

## Year2006          0.2971      0.1506      1.97      0.0491 *
## Year2007          0.3979      0.1614      2.46      0.0141 *
## Year2008          0.1434      0.1590      0.90      0.3677
## Year2009          0.1616      0.1636      0.99      0.3238
## Year2010          0.0780      0.1755      0.44      0.6571
## Year2011          0.2890      0.1570      1.84      0.0663 .
## Year2012          0.2263      0.1496      1.51      0.1311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.0759, Adjusted R-squared:  0.0347
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 390 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.263  0.872  0.954  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      2.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.149 1      1.072
## Year              1.149 16      1.004

```

Residuals from first author



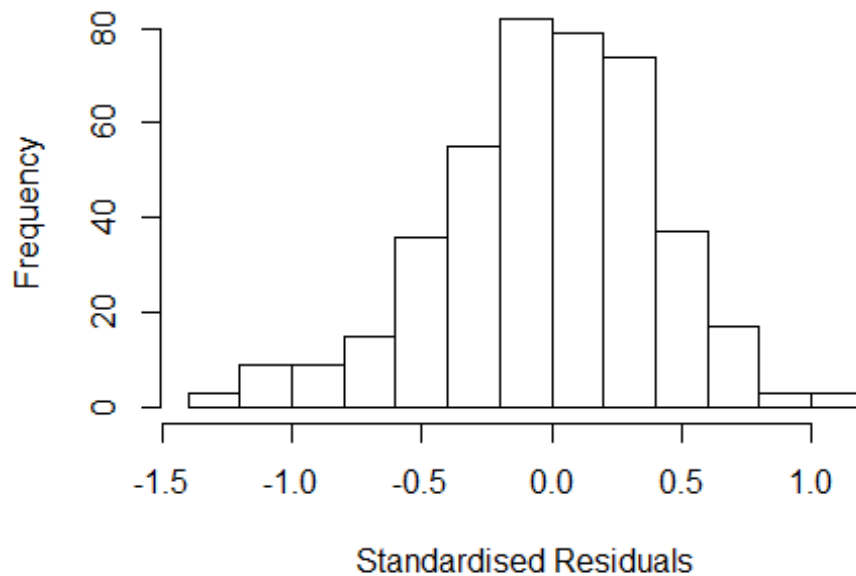
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26119 -0.25548  0.00648  0.25527  1.13271
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8632    0.1347    6.41 4.1e-10 ***
## FirstAuthorFemale1 -0.0608    0.0458   -1.33  0.1850
## Year1997         0.1143    0.2043    0.56  0.5761
## Year1998         0.4910    0.1770    2.77  0.0058 **
## Year1999         0.2742    0.1659    1.65  0.0990 .
## Year2000         0.4733    0.1481    3.20  0.0015 **
## Year2001         0.3002    0.1695    1.77  0.0772 .
## Year2002         0.2774    0.1556    1.78  0.0753 .
## Year2003         0.3619    0.1486    2.44  0.0153 *
## Year2004         0.2668    0.1487    1.79  0.0734 .
## Year2005         0.3545    0.1587    2.23  0.0261 *
## Year2006         0.2985    0.1497    1.99  0.0469 *
```

```

## Year2007          0.3980      0.1609      2.47      0.0138 *
## Year2008          0.1441      0.1583      0.91      0.3633
## Year2009          0.1619      0.1633      0.99      0.3221
## Year2010          0.0781      0.1752      0.45      0.6562
## Year2011          0.2897      0.1563      1.85      0.0646 .
## Year2012          0.2271      0.1488      1.53      0.1277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0759, Adjusted R-squared:  0.037
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 390 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.873   0.952   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      2.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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

```

Residuals from last author



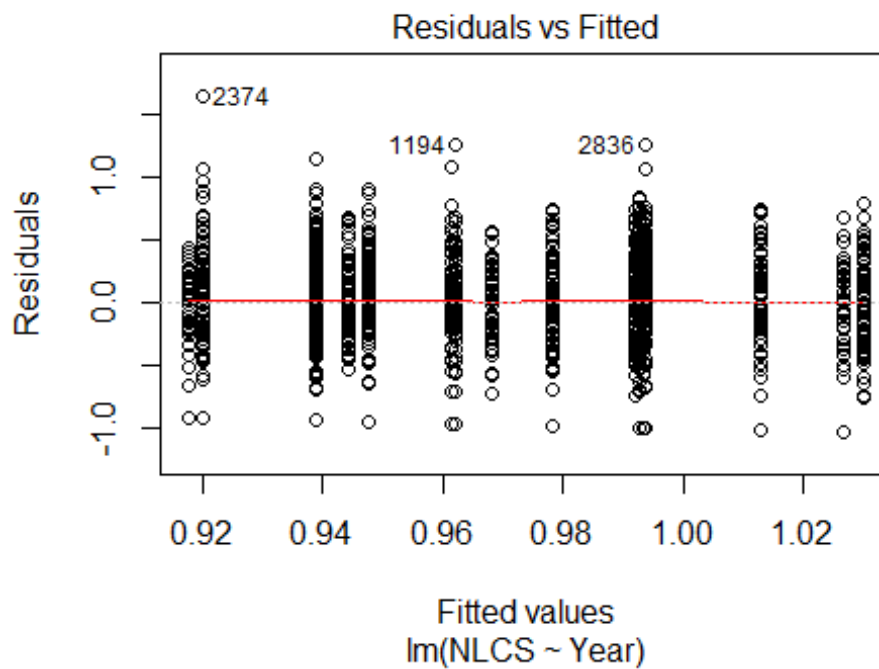
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.246331 -0.255772 -0.000813  0.257896  1.137553
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8502     0.1416   6.01 4.3e-09 ***
## LastAuthorFemale1 -0.0194     0.0474  -0.41  0.6825
## Year1997          0.1082     0.2096   0.52  0.6062
## Year1998          0.4839     0.1774   2.73  0.0066 **
## Year1999          0.2647     0.1672   1.58  0.1140
## Year2000          0.4726     0.1536   3.08  0.0022 **
## Year2001          0.3032     0.1754   1.73  0.0846 .
## Year2002          0.2788     0.1591   1.75  0.0805 .
## Year2003          0.3580     0.1532   2.34  0.0200 *
## Year2004          0.2762     0.1532   1.80  0.0721 .
## Year2005          0.3530     0.1631   2.16  0.0311 *
## Year2006          0.2920     0.1538   1.90  0.0583 .
```

```

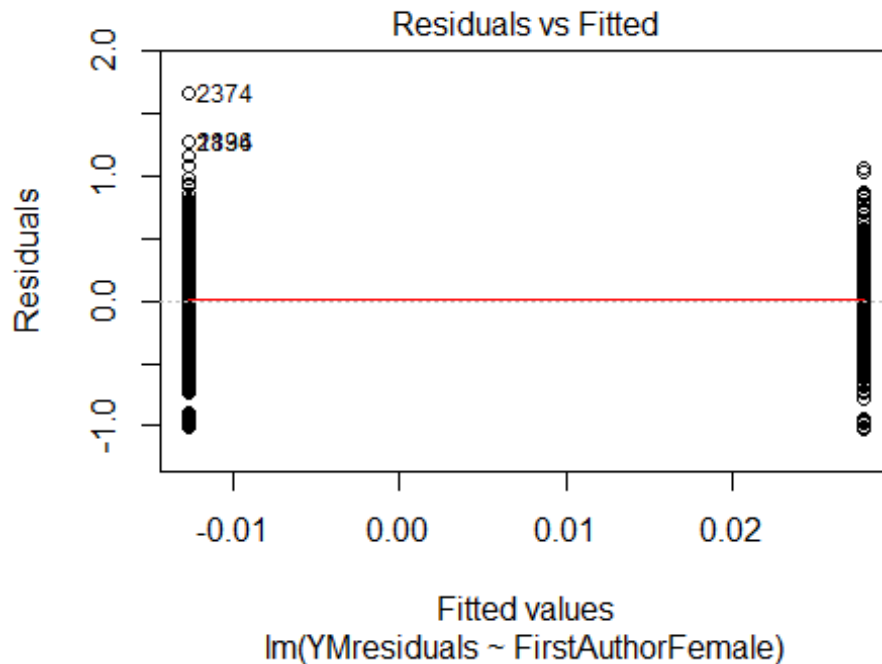
## Year2007          0.3962      0.1645      2.41      0.0165 *
## Year2008          0.1346      0.1628      0.83      0.4090
## Year2009          0.1649      0.1673      0.99      0.3249
## Year2010          0.0863      0.1772      0.49      0.6266
## Year2011          0.2837      0.1586      1.79      0.0745 .
## Year2012          0.2354      0.1531      1.54      0.1249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0717, Adjusted R-squared:  0.0326
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 395 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.274  0.871  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      2.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 422"
## [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
## 138 123 138 150 137 117 142 139 135 131 153 168 172 156 177
## 2011 2012
## 144 132
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 36 56 74 53 46 95 78 86 83 95 112 106 102 116
## 2011 2012

```

```
## 106 88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 34 50 67 51 39 77 65 75 71 84 103 94 95 104
## 2011 2012
## 97 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 = 58, df = 16, p-value = 1e-06
```



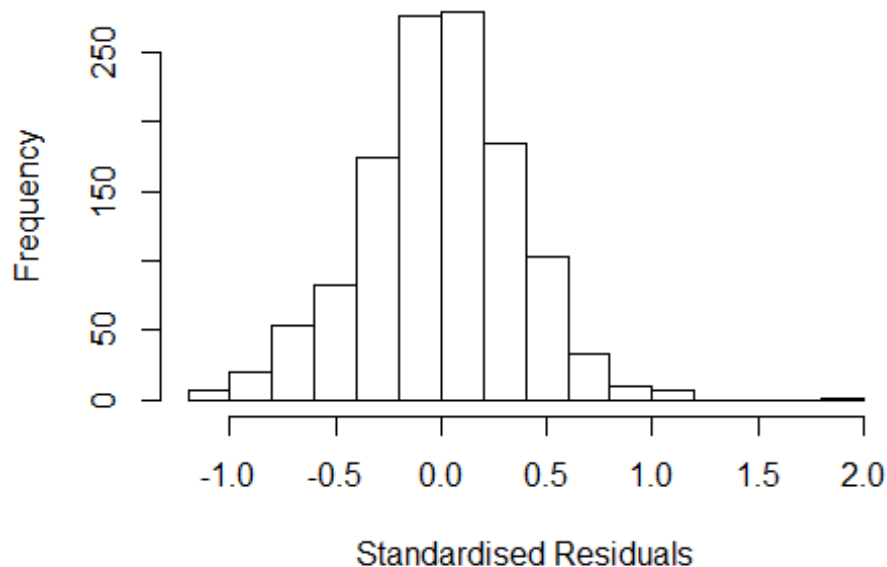
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.8, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 3.81749968109506"
## [1] "Male first author team size 2018 geometric mean: 2.60612817894307"
## 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 = 940, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.75336956588809"
## [1] "Male last author team size 2018 geometric mean: 2.99480919297277"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 350, 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	1.168	1	1.081
LastAuthorFemale	1.157	1	1.076
UniqueAuthors	1.368	4	1.040
Year	1.569	16	1.014

Residuals from first and last author and team size



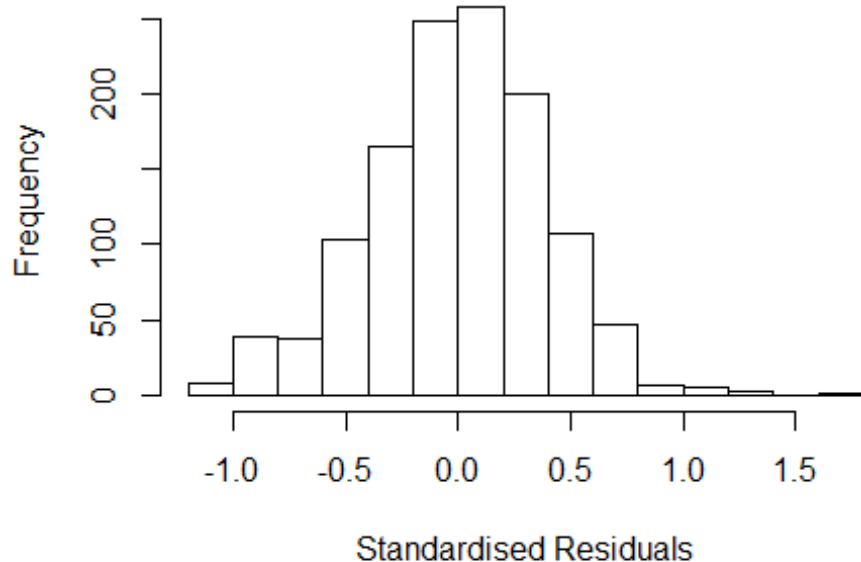
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1151 -0.2229 0.0035 0.2273 1.9054
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.802639 0.065809 12.20 < 2e-16 ***
## FirstAuthorFemale1 0.000504 0.024295 0.02 0.983
## LastAuthorFemale1 -0.031152 0.027535 -1.13 0.258
## UniqueAuthors2 0.227331 0.031466 7.22 8.9e-13 ***
## UniqueAuthors3 0.278930 0.033088 8.43 < 2e-16 ***
## UniqueAuthors4 0.364671 0.039245 9.29 < 2e-16 ***
## UniqueAuthors5 0.420305 0.034133 12.31 < 2e-16 ***
## Year1997 -0.012696 0.085724 -0.15 0.882
## Year1998 0.000729 0.080358 0.01 0.993
## Year1999 -0.042503 0.072151 -0.59 0.556
```

```

## Year2000      -0.046494    0.077413   -0.60    0.548
## Year2001      -0.075272    0.077561   -0.97    0.332
## Year2002      -0.092148    0.076041   -1.21    0.226
## Year2003      -0.051040    0.074928   -0.68    0.496
## Year2004      -0.039280    0.072237   -0.54    0.587
## Year2005       0.008275    0.078142    0.11    0.916
## Year2006      -0.024221    0.074032   -0.33    0.744
## Year2007      -0.081270    0.072674   -1.12    0.264
## Year2008      -0.052218    0.072258   -0.72    0.470
## Year2009      -0.138011    0.076567   -1.80    0.072 .
## Year2010      -0.076666    0.073070   -1.05    0.294
## Year2011      -0.048448    0.079805   -0.61    0.544
## Year2012      -0.113635    0.085072   -1.34    0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared:  0.146, Adjusted R-squared:  0.13
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 906 is an outlier with |weight| = 0 ( < 8.1e-05);
## 108 weights are ~= 1. The remaining 1121 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.244  0.863   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           8.13e-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 1.121 1 1.059
## LastAuthorFemale 1.094 1 1.046
## Year 1.153 16 1.004

```

Residuals from first and last author



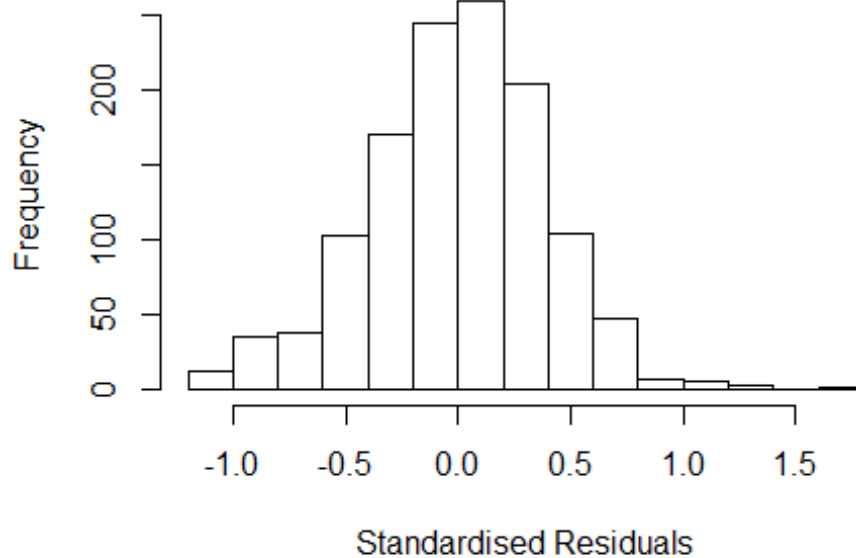
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0195 -0.2519 0.0084 0.2537 1.6723
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9989 0.0616 16.21 <2e-16 ***
## FirstAuthorFemale1 0.0397 0.0255 1.56 0.12
## LastAuthorFemale1 -0.0408 0.0302 -1.35 0.18
## Year1997 0.0205 0.0843 0.24 0.81
## Year1998 -0.0349 0.0840 -0.42 0.68
## Year1999 -0.0802 0.0743 -1.08 0.28
## Year2000 -0.0219 0.0790 -0.28 0.78
## Year2001 -0.0771 0.0735 -1.05 0.29
## Year2002 -0.0537 0.0743 -0.72 0.47
## Year2003 -0.0351 0.0753 -0.47 0.64
## Year2004 -0.0300 0.0735 -0.41 0.68
## Year2005 0.0144 0.0787 0.18 0.85
```

```

## Year2006          0.0150      0.0724      0.21      0.84
## Year2007          -0.0335     0.0719     -0.47     0.64
## Year2008          -0.0149     0.0728     -0.20     0.84
## Year2009          -0.1012     0.0788     -1.29     0.20
## Year2010          -0.0407     0.0735     -0.55     0.58
## Year2011           0.0188     0.0802      0.23     0.81
## Year2012          -0.0529     0.0869     -0.61     0.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0115, Adjusted R-squared:  -0.00321
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1124 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0096 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      8.13e-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 1.088 1      1.043
## Year              1.088 16      1.003

```

Residuals from first author



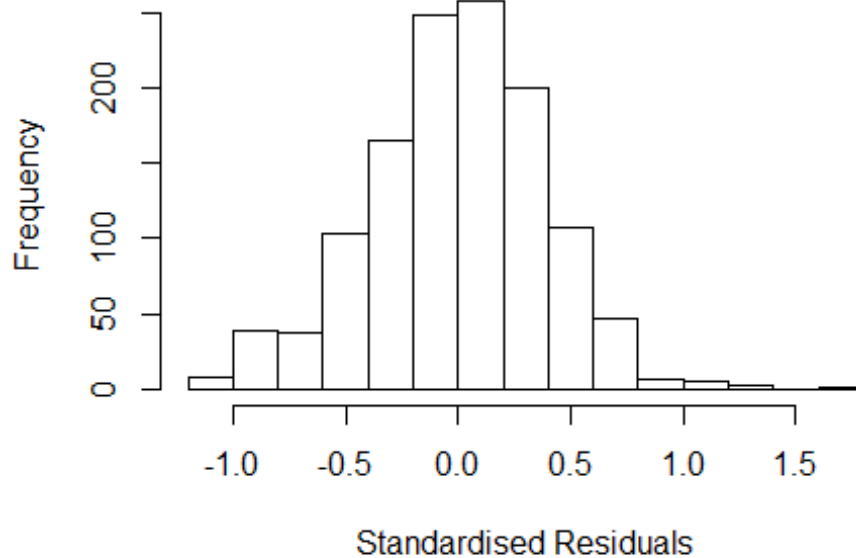
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0480 -0.2470 0.0116 0.2551 1.6772
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9902 0.0609 16.26 <2e-16 ***
## FirstAuthorFemale1 0.0334 0.0253 1.32 0.19
## Year1997 0.0214 0.0845 0.25 0.80
## Year1998 -0.0373 0.0837 -0.45 0.66
## Year1999 -0.0798 0.0740 -1.08 0.28
## Year2000 -0.0195 0.0788 -0.25 0.80
## Year2001 -0.0750 0.0727 -1.03 0.30
## Year2002 -0.0549 0.0740 -0.74 0.46
## Year2003 -0.0343 0.0747 -0.46 0.65
## Year2004 -0.0243 0.0726 -0.33 0.74
## Year2005 0.0166 0.0778 0.21 0.83
## Year2006 0.0199 0.0718 0.28 0.78
```

```

## Year2007          -0.0296      0.0716   -0.41      0.68
## Year2008          -0.0138      0.0725   -0.19      0.85
## Year2009          -0.0974      0.0784   -1.24      0.21
## Year2010          -0.0383      0.0731   -0.52      0.60
## Year2011           0.0244      0.0794     0.31      0.76
## Year2012          -0.0505      0.0863   -0.58      0.56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.00989,    Adjusted R-squared:  -0.004
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0092 0.8690 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      8.13e-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 1.062 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
## -1.03086 -0.24852  0.00766  0.25743  1.65642
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0057     0.0616   16.33  <2e-16 ***
## LastAuthorFemale1 -0.0327     0.0294   -1.11    0.27
## Year1997          0.0214     0.0837    0.26    0.80
## Year1998         -0.0322     0.0836   -0.39    0.70
## Year1999         -0.0787     0.0740   -1.06    0.29
## Year2000         -0.0223     0.0795   -0.28    0.78
## Year2001         -0.0779     0.0736   -1.06    0.29
## Year2002         -0.0524     0.0744   -0.70    0.48
## Year2003         -0.0276     0.0747   -0.37    0.71
## Year2004         -0.0284     0.0735   -0.39    0.70
## Year2005          0.0181     0.0783    0.23    0.82
## Year2006          0.0165     0.0725    0.23    0.82
```

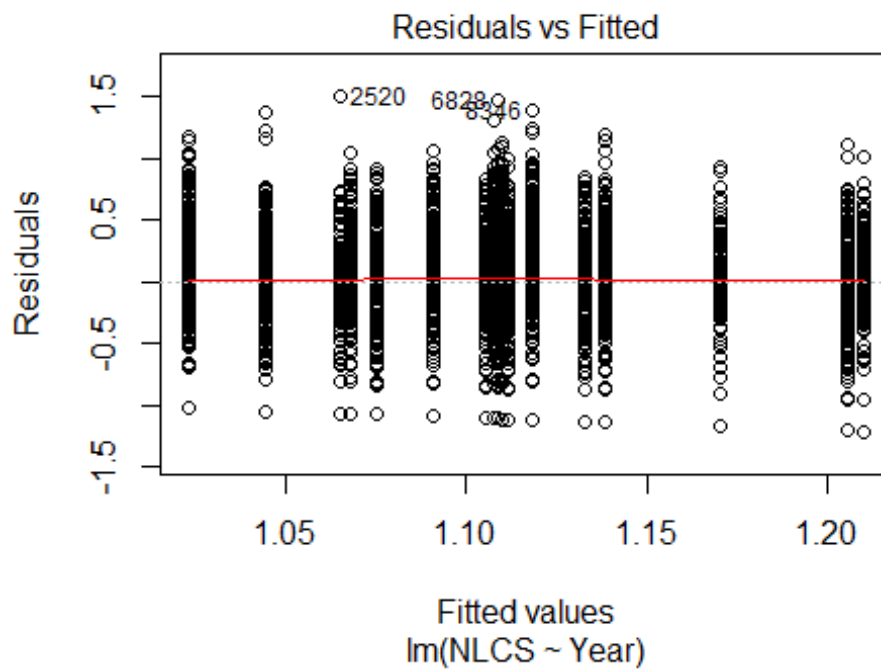
```

## Year2007          -0.0265      0.0719   -0.37      0.71
## Year2008          -0.0109      0.0727   -0.15      0.88
## Year2009          -0.0921      0.0788   -1.17      0.24
## Year2010          -0.0361      0.0734   -0.49      0.62
## Year2011           0.0252      0.0799    0.32      0.75
## Year2012          -0.0500      0.0868   -0.58      0.57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.00935,    Adjusted R-squared:  -0.00454
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0161 0.8700 0.9520 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      8.13e-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: 1230"
## [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
## 454 413 401 371 370 424 397 401 397 418 482 398 358 527 525
## 2011 2012
## 493 452
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 151 135 176 164 115 120 219 221 194 244 275 241 202 298 299
## 2011 2012

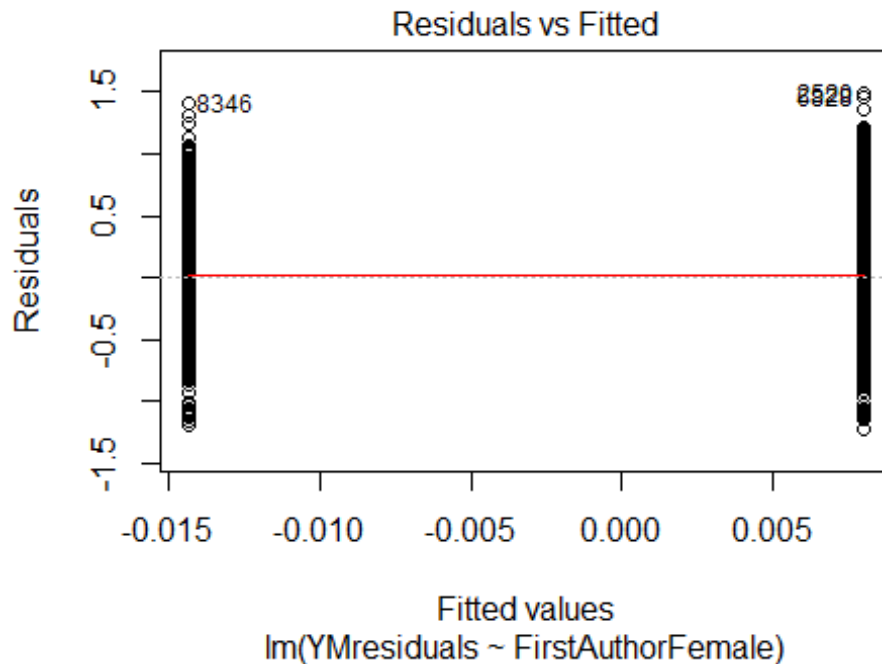
```



```
## 322 287
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 132 118 161 149 102 107 188 200 175 227 260 223 180 267 267
## 2011 2012
## 287 254
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 85, df = 16, p-value = 2e-11
```

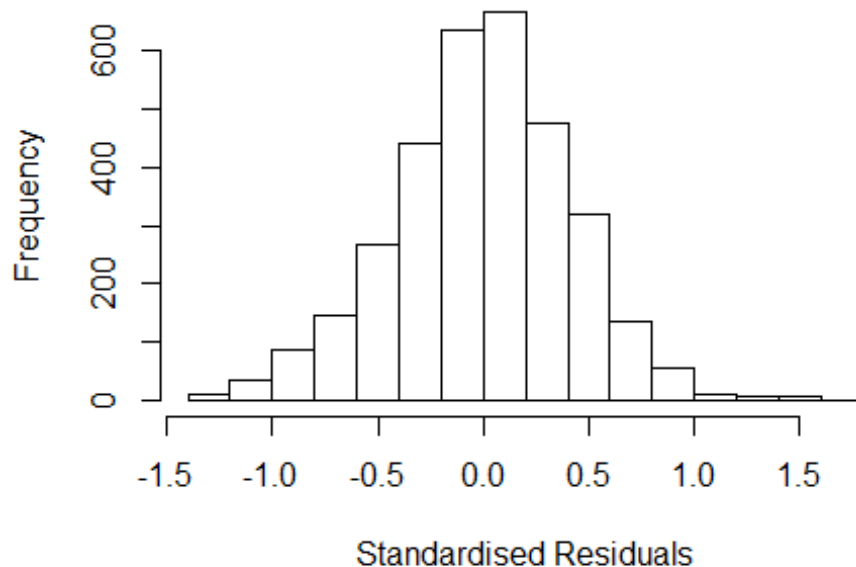


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.1, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.53326991542976"
## [1] "Male first author team size 2018 geometric mean: 3.43119480451977"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6200, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.57707183875412"
## [1] "Male last author team size 2018 geometric mean: 3.43339047203095"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5100, 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 1.047 1 1.023
## LastAuthorFemale 1.068 1 1.034
## UniqueAuthors 1.196 4 1.023
## Year 1.259 16 1.007
```

Residuals from first and last author and team size



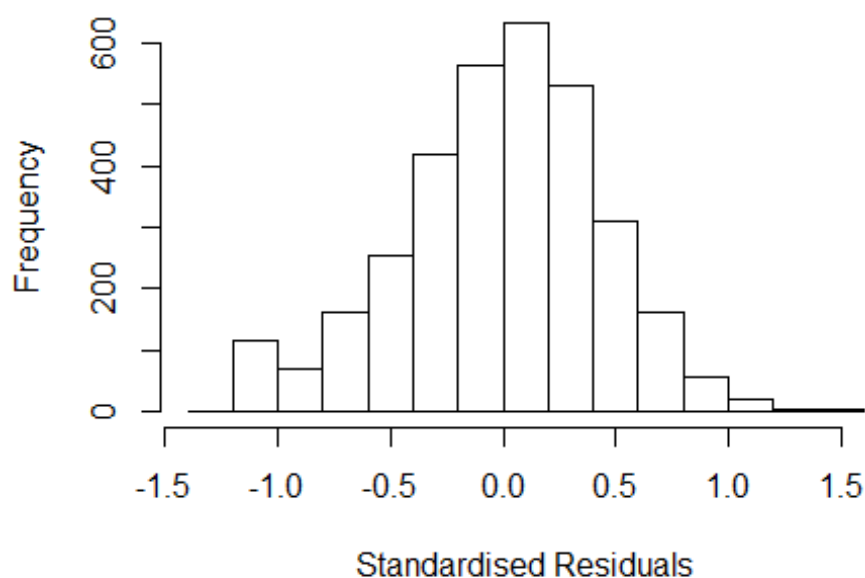
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31236 -0.25858  0.00925  0.27033  1.72295
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0396    0.0401   25.89 < 2e-16 ***
## FirstAuthorFemale1 -0.0419    0.0153   -2.73  0.00628 **
## LastAuthorFemale1 -0.0725    0.0178   -4.08  4.6e-05 ***
## UniqueAuthors2     0.2136    0.0283    7.54  5.9e-14 ***
## UniqueAuthors3     0.3169    0.0275   11.53 < 2e-16 ***
## UniqueAuthors4     0.4226    0.0292   14.47 < 2e-16 ***
## UniqueAuthors5     0.5232    0.0288   18.15 < 2e-16 ***
## Year1997          -0.0115    0.0499   -0.23  0.81829
## Year1998          -0.1719    0.0494   -3.48  0.00051 ***
## Year1999          -0.1212    0.0472   -2.57  0.01029 *
```

```

## Year2000          -0.1083      0.0498   -2.17   0.02982 *
## Year2001          -0.2012      0.0489   -4.11   4.0e-05 ***
## Year2002          -0.1870      0.0445   -4.21   2.7e-05 ***
## Year2003          -0.2036      0.0451   -4.51   6.7e-06 ***
## Year2004          -0.2480      0.0449   -5.52   3.6e-08 ***
## Year2005          -0.2593      0.0437   -5.93   3.3e-09 ***
## Year2006          -0.1893      0.0435   -4.35   1.4e-05 ***
## Year2007          -0.2504      0.0474   -5.28   1.4e-07 ***
## Year2008          -0.2352      0.0488   -4.82   1.5e-06 ***
## Year2009          -0.1925      0.0467   -4.12   3.9e-05 ***
## Year2010          -0.2067      0.0457   -4.52   6.4e-06 ***
## Year2011          -0.2145      0.0455   -4.72   2.5e-06 ***
## Year2012          -0.2981      0.0481   -6.20   6.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.15, Adjusted R-squared:  0.144
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 283 weights are ~= 1. The remaining 3014 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0216 0.8630 0.9500 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          3.03e-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 1.042 1 1.021
## LastAuthorFemale 1.054 1 1.027
## Year 1.075 16 1.002

```

Residuals from first and last author



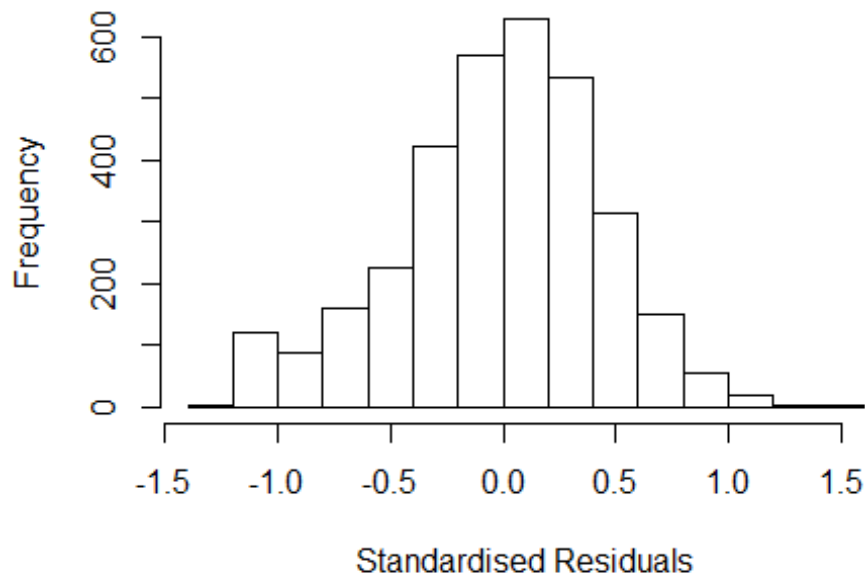
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2634 -0.2782  0.0231  0.2801  1.5588
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2463    0.0367   33.94 < 2e-16 ***
## FirstAuthorFemale1 -0.0121    0.0162   -0.75  0.45623
## LastAuthorFemale1  -0.0961    0.0192   -5.02  5.6e-07 ***
## Year1997           0.0171    0.0505    0.34  0.73491
## Year1998          -0.1536    0.0495   -3.10  0.00194 **
## Year1999          -0.0935    0.0482   -1.94  0.05237 .
## Year2000          -0.0691    0.0511   -1.35  0.17693
## Year2001          -0.1470    0.0506   -2.91  0.00366 **
## Year2002          -0.0894    0.0453   -1.97  0.04850 *
## Year2003          -0.1124    0.0461   -2.44  0.01471 *
## Year2004          -0.1514    0.0467   -3.24  0.00121 **
## Year2005          -0.1496    0.0447   -3.35  0.00082 ***
```

```

## Year2006          -0.0800      0.0450   -1.78  0.07596 .
## Year2007          -0.1154      0.0487   -2.37  0.01779 *
## Year2008          -0.1084      0.0506   -2.14  0.03214 *
## Year2009          -0.0874      0.0481   -1.82  0.06919 .
## Year2010          -0.0774      0.0481   -1.61  0.10755 .
## Year2011          -0.0914      0.0472   -1.94  0.05277 .
## Year2012          -0.1811      0.0518   -3.50  0.00047 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.0137
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 252 weights are ~= 1. The remaining 3045 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.142  0.864  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      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 1.025 1      1.012
## Year              1.025 16      1.001

```

Residuals from first author



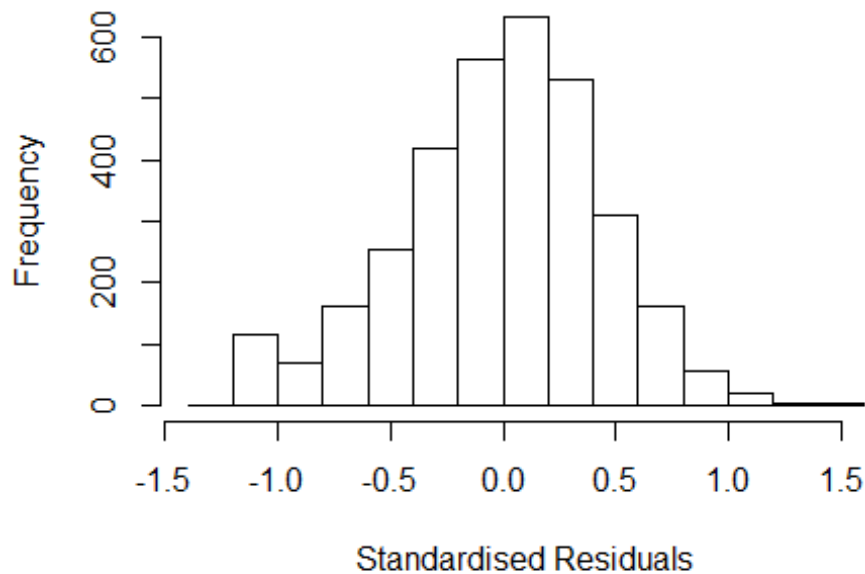
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2460 -0.2798 0.0248 0.2822 1.4706
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2318 0.0365 33.78 < 2e-16 ***
## FirstAuthorFemale1 -0.0261 0.0163 -1.60 0.10959
## Year1997 0.0142 0.0507 0.28 0.77895
## Year1998 -0.1615 0.0490 -3.30 0.00099 ***
## Year1999 -0.0988 0.0482 -2.05 0.04019 *
## Year2000 -0.0735 0.0510 -1.44 0.14956
## Year2001 -0.1403 0.0506 -2.77 0.00561 **
## Year2002 -0.0897 0.0453 -1.98 0.04790 *
## Year2003 -0.1095 0.0461 -2.38 0.01758 *
## Year2004 -0.1545 0.0468 -3.30 0.00098 ***
## Year2005 -0.1489 0.0447 -3.33 0.00089 ***
## Year2006 -0.0801 0.0452 -1.77 0.07679 .
```

```

## Year2007          -0.1159      0.0484   -2.39  0.01678 *
## Year2008          -0.1142      0.0503   -2.27  0.02319 *
## Year2009          -0.0947      0.0482   -1.96  0.04964 *
## Year2010          -0.0811      0.0485   -1.67  0.09482 .
## Year2011          -0.0950      0.0472   -2.01  0.04425 *
## Year2012          -0.1757      0.0514   -3.42  0.00063 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0111, Adjusted R-squared:  0.00593
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 241 weights are ~= 1. The remaining 3056 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.201  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      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 1.039 1      1.019
## Year      1.039 16      1.001

```


Residuals from last author



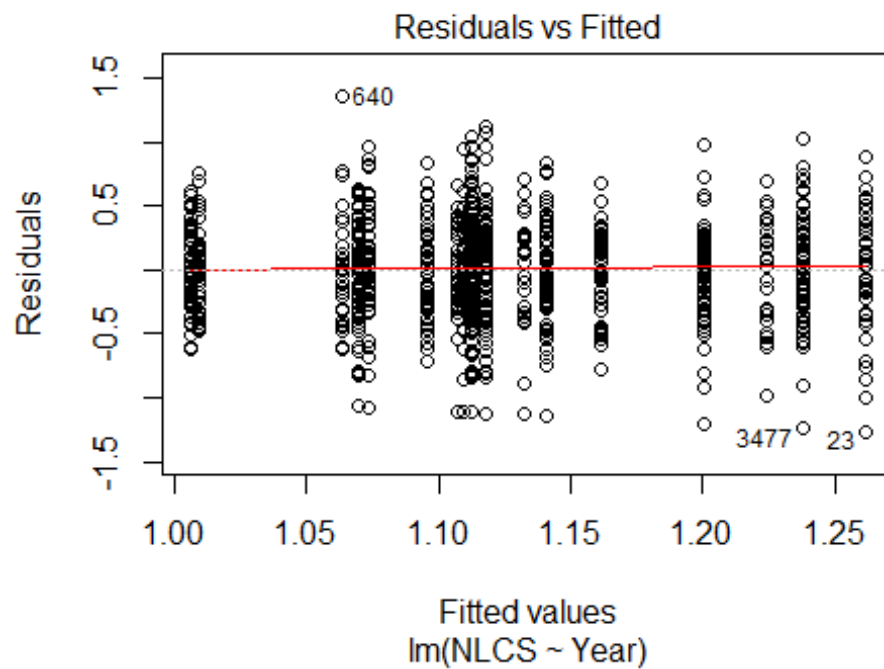
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2605 -0.2817 0.0232 0.2791 1.5653
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2424 0.0363 34.27 < 2e-16 ***
## LastAuthorFemale1 -0.0985 0.0191 -5.15 2.7e-07 ***
## Year1997 0.0181 0.0505 0.36 0.71982
## Year1998 -0.1532 0.0495 -3.10 0.00198 **
## Year1999 -0.0927 0.0482 -1.93 0.05422 .
## Year2000 -0.0683 0.0511 -1.34 0.18139
## Year2001 -0.1472 0.0506 -2.91 0.00362 **
## Year2002 -0.0900 0.0453 -1.99 0.04699 *
## Year2003 -0.1137 0.0460 -2.47 0.01351 *
## Year2004 -0.1512 0.0467 -3.24 0.00123 **
## Year2005 -0.1496 0.0447 -3.35 0.00082 ***
## Year2006 -0.0797 0.0450 -1.77 0.07683 .
```

```

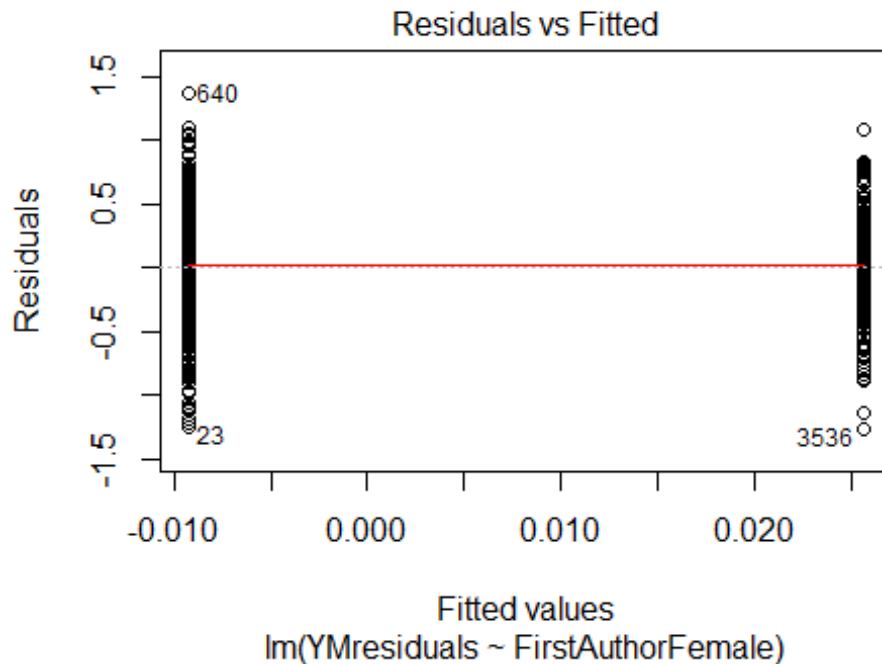
## Year2007          -0.1151      0.0487   -2.36  0.01824 *
## Year2008          -0.1081      0.0505   -2.14  0.03247 *
## Year2009          -0.0875      0.0481   -1.82  0.06881 .
## Year2010          -0.0774      0.0481   -1.61  0.10760
## Year2011          -0.0916      0.0471   -1.94  0.05219 .
## Year2012          -0.1816      0.0517   -3.51  0.00045 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0189, Adjusted R-squared:  0.0138
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 251 weights are ~ = 1. The remaining 3046 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.865  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      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: 3297"
## [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
##  187  204  181  152  151  164  168  161  139  130  186  172  154  150  184
## 2011 2012
##  217  194
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   44   33   32   36   28   27   57   49   45   47   68   70   56   59   80
## 2011 2012

```

```
## 116 86
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 27 26 30 22 26 34 45 37 36 58 59 46 55 65
## 2011 2012
## 101 76
## [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
```

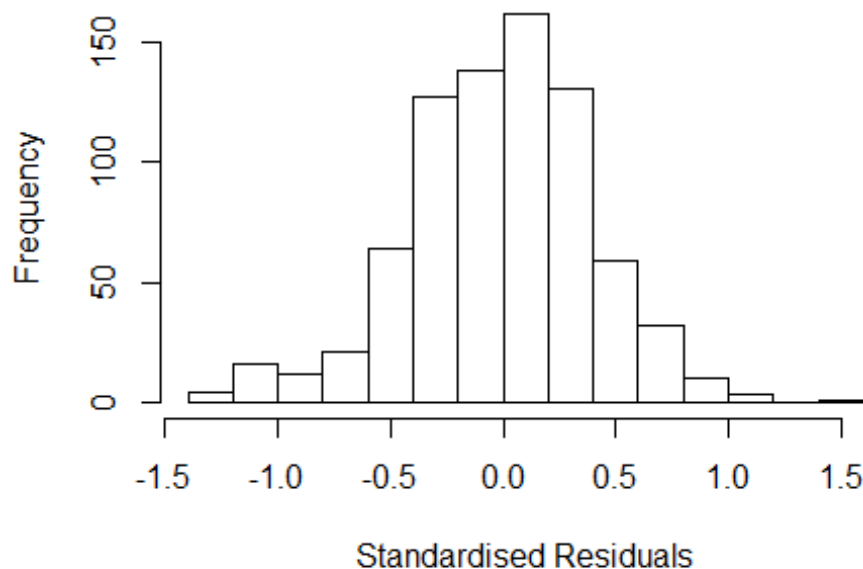


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 3.61057445774624"
## [1] "Male first author team size 2018 geometric mean: 3.3913339419382"
##
## 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: 3.95274856346309"
## [1] "Male last author team size 2018 geometric mean: 3.35196648046027"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, 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 1.157 1      1.076
## LastAuthorFemale  1.162 1      1.078
## UniqueAuthors    1.487 4      1.051
## Year              1.669 16     1.016
```

Residuals from first and last author and team size



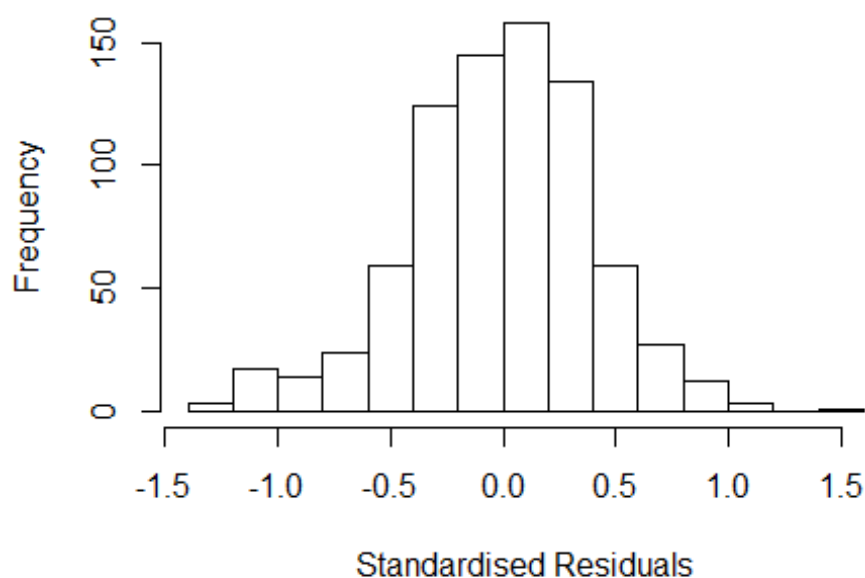
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3059 -0.2584  0.0102  0.2563  1.4485
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.27259    0.08928   14.25 < 2e-16 ***
## FirstAuthorFemale1  0.02624    0.03188    0.82  0.41075
## LastAuthorFemale1  0.04504    0.03622    1.24  0.21414
## UniqueAuthors2   -0.00347    0.05242   -0.07  0.94720
## UniqueAuthors3    0.09045    0.05193    1.74  0.08192 .
## UniqueAuthors4    0.06799    0.05538    1.23  0.21993
## UniqueAuthors5    0.08610    0.06686    1.29  0.19820
## Year1997         -0.21458    0.10410   -2.06  0.03962 *
## Year1998         -0.30313    0.10619   -2.85  0.00443 **
## Year1999         -0.18853    0.11070   -1.70  0.08896 .
```

```

## Year2000      -0.09891    0.13117   -0.75  0.45104
## Year2001      -0.13644    0.11236   -1.21  0.22499
## Year2002      -0.31563    0.10430   -3.03  0.00256 **
## Year2003      -0.25112    0.09887   -2.54  0.01129 *
## Year2004      -0.35903    0.09445   -3.80  0.00016 ***
## Year2005      -0.19196    0.09525   -2.02  0.04423 *
## Year2006      -0.24672    0.09792   -2.52  0.01196 *
## Year2007      -0.15131    0.09163   -1.65  0.09909 .
## Year2008      -0.16020    0.09697   -1.65  0.09896 .
## Year2009      -0.24611    0.10354   -2.38  0.01771 *
## Year2010      -0.23211    0.10095   -2.30  0.02176 *
## Year2011      -0.23173    0.09144   -2.53  0.01147 *
## Year2012      -0.08342    0.09250   -0.90  0.36742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0548, Adjusted R-squared:  0.0274
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 719 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.875  0.951  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          1.28e-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 1.133 1          1.064
## LastAuthorFemale 1.127 1          1.061
## Year 1.225 16          1.006

```

Residuals from first and last author



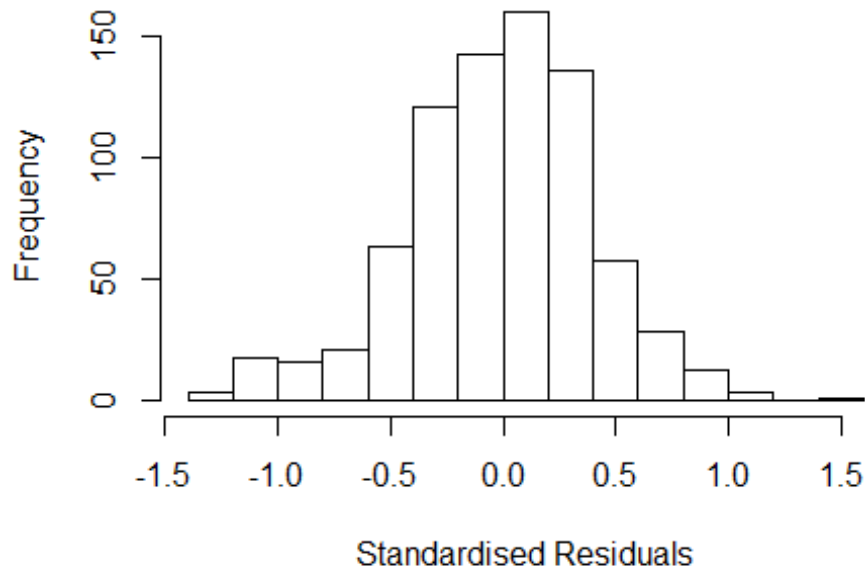
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29170 -0.26325  0.00417  0.25070  1.41884
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2917    0.0813   15.90 < 2e-16 ***
## FirstAuthorFemale1  0.0346    0.0322    1.07  0.28341
## LastAuthorFemale1  0.0482    0.0358    1.35  0.17888
## Year1997         -0.2167    0.1032   -2.10  0.03613 *
## Year1998         -0.2925    0.1095   -2.67  0.00773 **
## Year1999         -0.1603    0.1081   -1.48  0.13825
## Year2000         -0.0798    0.1269   -0.63  0.52942
## Year2001         -0.1153    0.1103   -1.05  0.29620
## Year2002         -0.2861    0.1031   -2.77  0.00566 **
## Year2003         -0.2326    0.0995   -2.34  0.01970 *
## Year2004         -0.3307    0.0950   -3.48  0.00053 ***
## Year2005         -0.1686    0.0960   -1.76  0.07953 .
```

```

## Year2006          -0.2139      0.0977   -2.19  0.02882 *
## Year2007          -0.1249      0.0906   -1.38  0.16869
## Year2008          -0.1330      0.0983   -1.35  0.17642
## Year2009          -0.2126      0.1020   -2.08  0.03748 *
## Year2010          -0.2065      0.1006   -2.05  0.04043 *
## Year2011          -0.2053      0.0914   -2.25  0.02496 *
## Year2012          -0.0503      0.0918   -0.55  0.58424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.0441, Adjusted R-squared:  0.0215
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 717 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.151  0.882  0.952  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      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 1.11 1          1.054
## Year              1.11 16          1.003

```


Residuals from first author



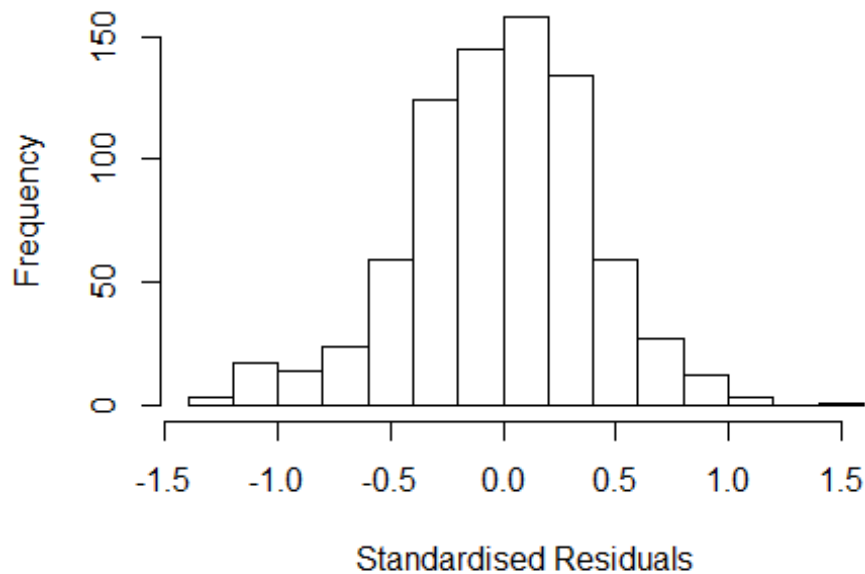
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29817 -0.26391 0.00773 0.25834 1.41591
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2982 0.0816 15.90 < 2e-16 ***
## FirstAuthorFemale1 0.0388 0.0320 1.21 0.22541
## Year1997 -0.2213 0.1034 -2.14 0.03266 *
## Year1998 -0.2961 0.1098 -2.70 0.00714 **
## Year1999 -0.1607 0.1089 -1.48 0.14035
## Year2000 -0.0761 0.1277 -0.60 0.55170
## Year2001 -0.1197 0.1107 -1.08 0.27981
## Year2002 -0.2866 0.1041 -2.75 0.00603 **
## Year2003 -0.2373 0.0997 -2.38 0.01754 *
## Year2004 -0.3312 0.0958 -3.46 0.00058 ***
## Year2005 -0.1711 0.0969 -1.77 0.07785 .
## Year2006 -0.2128 0.0983 -2.16 0.03073 *
```

```

## Year2007          -0.1219      0.0916   -1.33  0.18363
## Year2008          -0.1321      0.0987   -1.34  0.18125
## Year2009          -0.2142      0.1022   -2.10  0.03647 *
## Year2010          -0.2096      0.1013   -2.07  0.03887 *
## Year2011          -0.2042      0.0922   -2.22  0.02706 *
## Year2012          -0.0468      0.0925   -0.51  0.61303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.0426, Adjusted R-squared:  0.0213
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 720 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.879  0.953  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      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 1.102 1      1.050
## Year      1.102 16      1.003

```

Residuals from last author



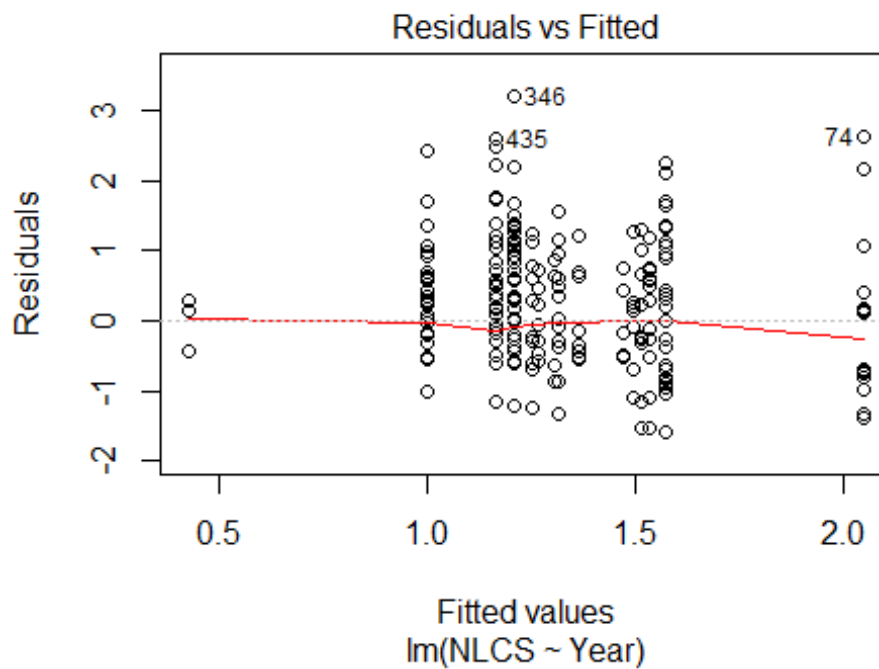
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29654 -0.26446  0.00809  0.24899  1.41549
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2965     0.0804   16.13  <2e-16 ***
## LastAuthorFemale1  0.0528     0.0354    1.49   0.1364
## Year1997        -0.2201     0.1027   -2.14   0.0324 *
## Year1998        -0.2940     0.1084   -2.71   0.0068 **
## Year1999        -0.1549     0.1080   -1.43   0.1520
## Year2000        -0.0778     0.1259   -0.62   0.5368
## Year2001        -0.1157     0.1097   -1.05   0.2923
## Year2002        -0.2869     0.1033   -2.78   0.0056 **
## Year2003        -0.2259     0.0994   -2.27   0.0233 *
## Year2004        -0.3265     0.0947   -3.45   0.0006 ***
## Year2005        -0.1660     0.0960   -1.73   0.0841 .
## Year2006        -0.2083     0.0974   -2.14   0.0328 *
```

```

## Year2007          -0.1187      0.0903   -1.31    0.1889
## Year2008          -0.1285      0.0981   -1.31    0.1905
## Year2009          -0.2064      0.1015   -2.03    0.0425 *
## Year2010          -0.1967      0.0996   -1.97    0.0486 *
## Year2011          -0.1993      0.0914   -2.18    0.0295 *
## Year2012          -0.0473      0.0916   -0.52    0.6059
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.0428, Adjusted R-squared:  0.0215
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 720 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.153  0.878  0.953  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      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: 780"
## [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
##    9    9   10    9    6   14   17   11   14   18   19   30   28   22   52
## 2011 2012
##   88   67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    5    7    4    3   10   14    9   12   14   14   21   17   14   36
## 2011 2012

```

```
## 68 52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 4 7 4 3 10 14 9 12 14 14 21 17 14 33
## 2011 2012
## 65 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 = 19, df = 16, p-value = 0.3
```



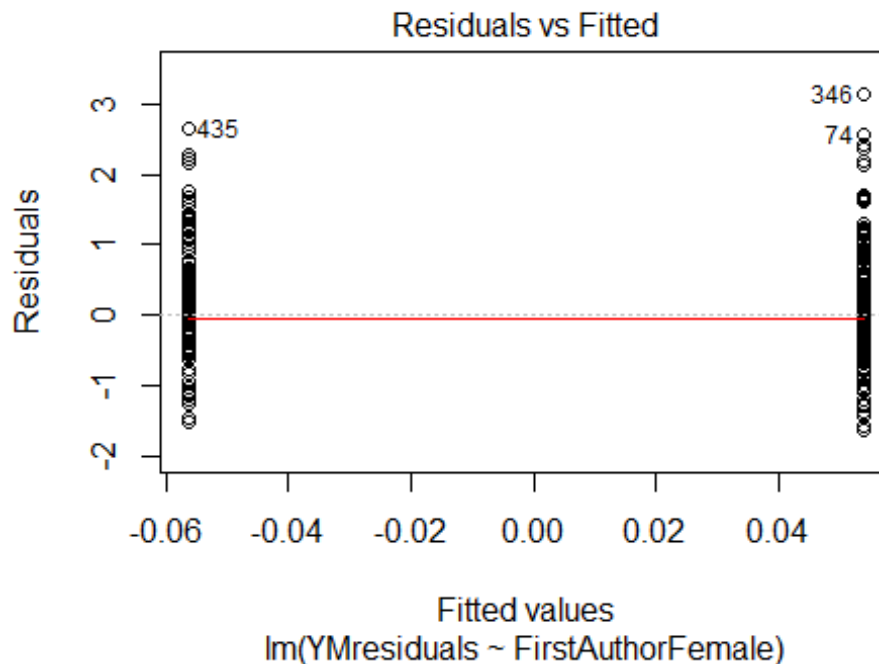
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.92, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 1.25992104989487"
## [1] "Male first author team size 2018 geometric mean: 1.44051951447658"

## 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 = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.55724668367001"
## [1] "Male last author team size 2018 geometric mean: 1.24469258946402"

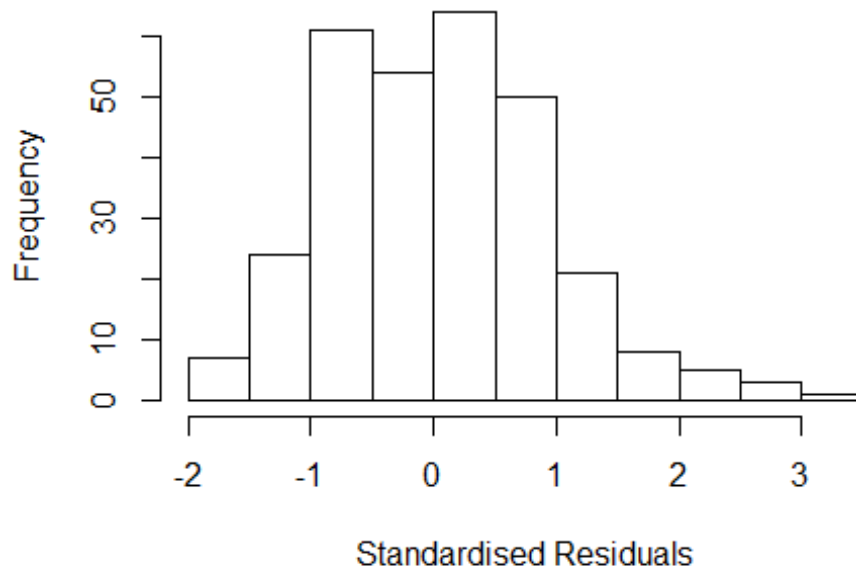
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	2.979	1	1.726
LastAuthorFemale	2.981	1	1.727
UniqueAuthors	2.869	4	1.141
Year	4.568	16	1.049

Residuals from first and last author and team size



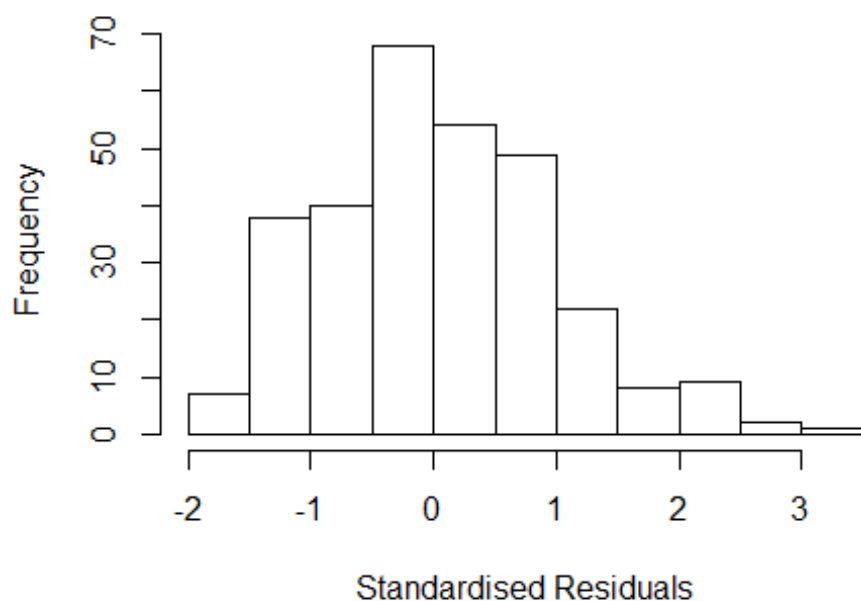
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 72   0141456147 4.203 2002    1200     1    2.559
## 74   0141791010 4.665 2002    1200     1    2.817
## 346  79952817615 4.404 2011    1200     2    3.295
## 435  84871086462 3.772 2012    1200     2    2.972
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7673 -0.6803  0.0171  0.6137  3.2946
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.34615    0.32512   4.14 4.6e-05 ***
## FirstAuthorFemale1  0.07256    0.17012   0.43 0.67008
## LastAuthorFemale1 -0.27732    0.16821  -1.65 0.10037
## UniqueAuthors2     0.55754    0.18323   3.04 0.00257 **
## UniqueAuthors3     0.57638    0.25068   2.30 0.02224 *
## UniqueAuthors4     0.58544    0.29085   2.01 0.04511 *
## UniqueAuthors5     0.53489    0.15434   3.47 0.00061 ***
```

```

## Year1997      -0.02870    0.40417   -0.07  0.94344
## Year1998      -0.06340    0.37403   -0.17  0.86552
## Year1999      -0.04158    0.53756   -0.08  0.93839
## Year2000      -1.12673    0.42610   -2.64  0.00866 **
## Year2001      -0.00395    0.39415   -0.01  0.99201
## Year2002       0.50223    0.43075    1.17  0.24465
## Year2003       0.13950    0.39171    0.36  0.72202
## Year2004      -0.16322    0.38503   -0.42  0.67196
## Year2005       0.17174    0.38456    0.45  0.65553
## Year2006       0.10738    0.47100    0.23  0.81983
## Year2007       0.12892    0.43567    0.30  0.76752
## Year2008      -0.32992    0.36325   -0.91  0.36454
## Year2009      -0.10776    0.40752   -0.26  0.79164
## Year2010      -0.37008    0.35507   -1.04  0.29821
## Year2011      -0.23679    0.35244   -0.67  0.50223
## Year2012      -0.34171    0.36271   -0.94  0.34697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.963
## Multiple R-squared:  0.143, Adjusted R-squared:  0.0748
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 270 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.218  0.903  0.953  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      3.36e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.842 1 1.960
## LastAuthorFemale 3.878 1 1.969
## Year 1.845 16 1.019

```


Residuals from first and last author



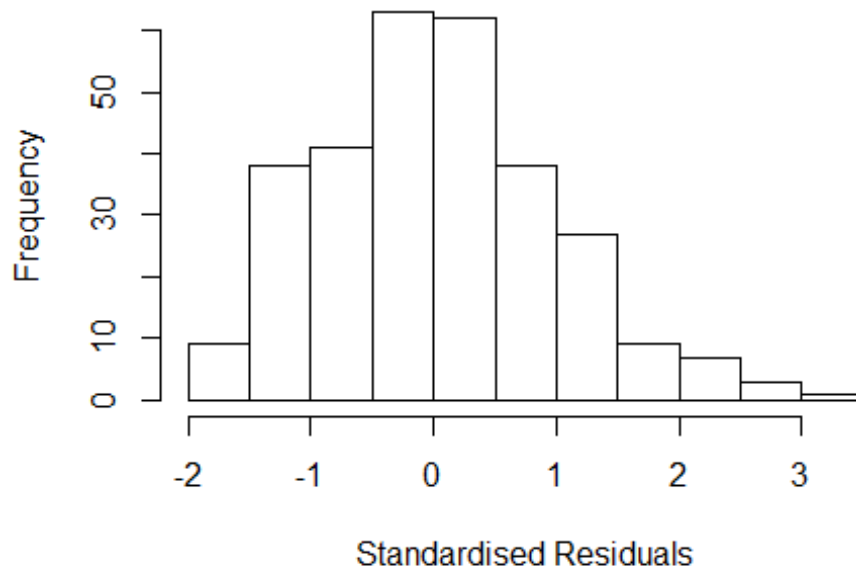
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 74   0141791010 4.665 2002    1200     1    2.733
## 346 79952817615 4.404 2011    1200     2    3.162
## 435 84871086462 3.772 2012    1200     2    2.799
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.674 -0.665 -0.050  0.617  3.162
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5106    0.2532   5.97 7.3e-09 ***
## FirstAuthorFemale1  0.2016    0.2043   0.99  0.325
## LastAuthorFemale1 -0.3921    0.2050  -1.91  0.057 .
## Year1997          0.2819    0.4092   0.69  0.492
## Year1998         -0.1282    0.3154  -0.41  0.685
## Year1999         -0.2060    0.4966  -0.41  0.679
## Year2000         -0.8915    0.3448  -2.59  0.010 *
## Year2001          0.0213    0.3688   0.06  0.954
## Year2002          0.4215    0.4026   1.05  0.296
## Year2003         -0.0344    0.3392  -0.10  0.919
```

```

## Year2004          -0.1894      0.3549   -0.53    0.594
## Year2005           0.1634      0.3527    0.46    0.643
## Year2006           0.0792      0.4753    0.17    0.868
## Year2007           0.0924      0.3985    0.23    0.817
## Year2008          -0.4026      0.3131   -1.29    0.200
## Year2009          -0.1496      0.3759   -0.40    0.691
## Year2010          -0.4732      0.2919   -1.62    0.106
## Year2011          -0.2689      0.2933   -0.92    0.360
## Year2012          -0.3469      0.3032   -1.14    0.254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.978
## Multiple R-squared:  0.0879, Adjusted R-squared:  0.029
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.275  0.897  0.955   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      3.36e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



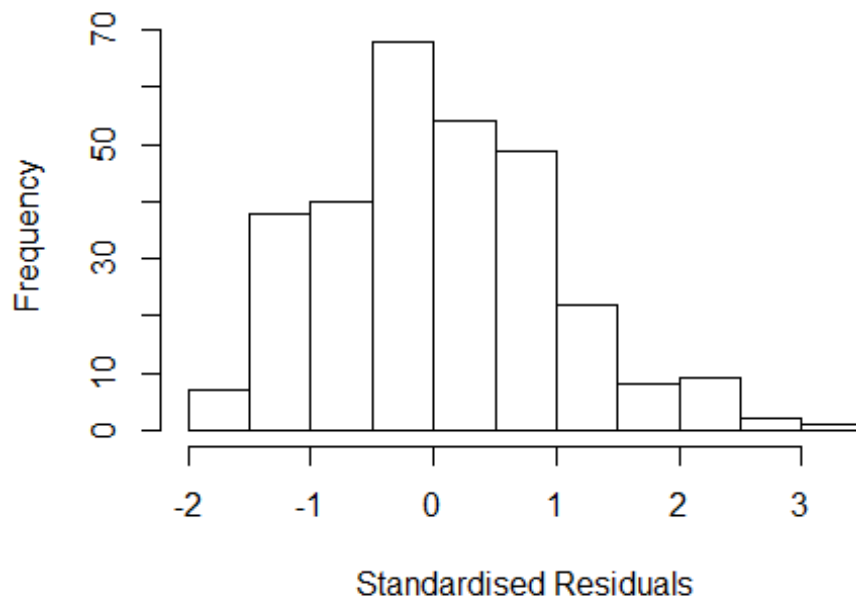
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 74   0141791010 4.665 2002    1200     1    2.733
## 346 79952817615 4.404 2011    1200     2    3.162
## 435 84871086462 3.772 2012    1200     2    2.799
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6422 -0.6408 -0.0161  0.5940  3.1954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4996    0.2608   5.75 2.3e-08 ***
## FirstAuthorFemale1 -0.1005    0.1199  -0.84  0.40292
## Year1997          0.0961    0.3663   0.26  0.79332
## Year1998         -0.2031    0.3164  -0.64  0.52140
## Year1999         -0.1950    0.5007  -0.39  0.69728
## Year2000         -1.0391    0.3041  -3.42  0.00073 ***
## Year2001          0.0812    0.3821   0.21  0.83180
## Year2002          0.4362    0.4238   1.03  0.30420
## Year2003         -0.0861    0.3488  -0.25  0.80527
## Year2004         -0.2098    0.3644  -0.58  0.56514
```

```

## Year2005          0.1426      0.3628      0.39  0.69454
## Year2006          0.0200      0.4668      0.04  0.96594
## Year2007          0.1177      0.3976      0.30  0.76739
## Year2008         -0.4578      0.3207     -1.43  0.15454
## Year2009         -0.1457      0.3732     -0.39  0.69647
## Year2010         -0.5080      0.3004     -1.69  0.09193 .
## Year2011         -0.2910      0.3006     -0.97  0.33395
## Year2012         -0.3724      0.3090     -1.20  0.22923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.976
## Multiple R-squared:  0.0739, Adjusted R-squared:  0.0177
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 278 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.886   0.956   0.915   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      3.36e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.329 1          1.153
## Year            1.329 16          1.009

```

Residuals from last author



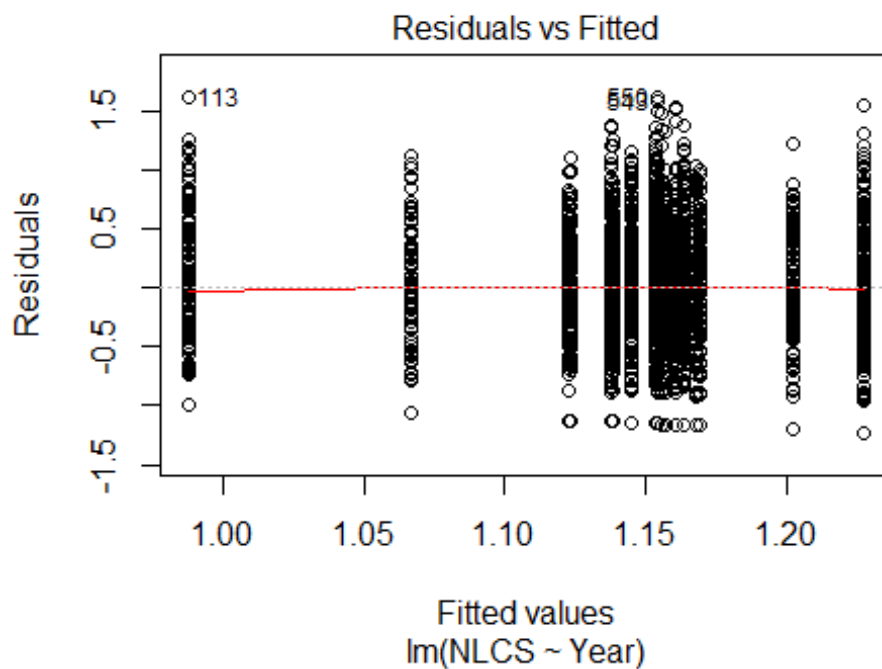
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 74   0141791010 4.665 2002    1200     1    2.733
## 346 79952817615 4.404 2011    1200     2    3.162
## 435 84871086462 3.772 2012    1200     2    2.799
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7188 -0.6767 -0.0444  0.5942  3.1263
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.51695    0.24911     6.09 3.7e-09 ***
## LastAuthorFemale1 -0.24104    0.12023    -2.00  0.0459 *
## Year1997          0.19969    0.38059     0.52  0.6002
## Year1998         -0.14228    0.31035    -0.46  0.6470
## Year1999         -0.21238    0.49477    -0.43  0.6681
## Year2000         -0.93052    0.30928    -3.01  0.0029 **
## Year2001          0.07635    0.35526     0.21  0.8300
## Year2002          0.47700    0.40332     1.18  0.2379
## Year2003         -0.00569    0.33228    -0.02  0.9863
## Year2004         -0.17796    0.34951    -0.51  0.6110
```

```

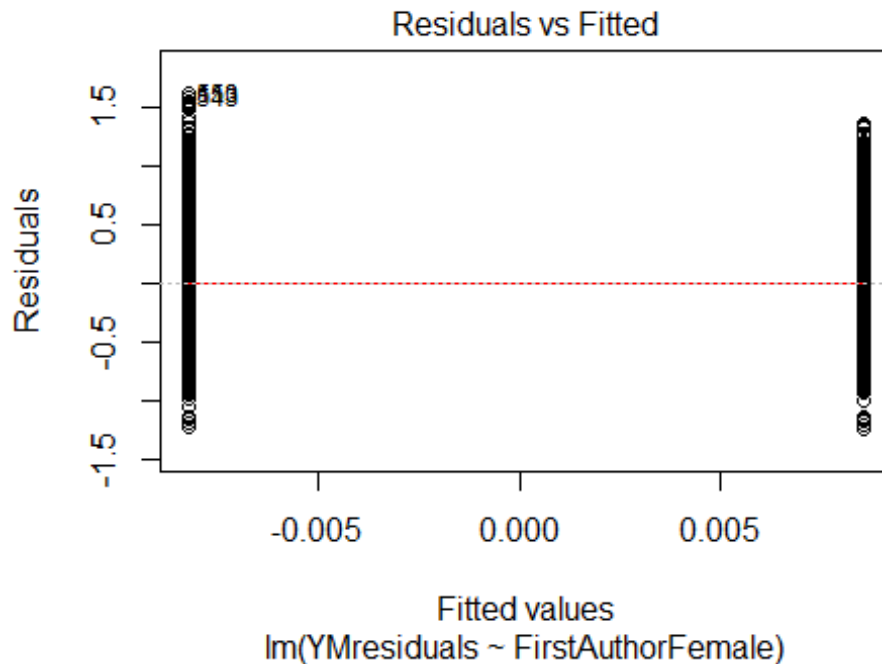
## Year2005      0.20187      0.35036      0.58      0.5650
## Year2006      0.06910      0.46963      0.15      0.8831
## Year2007      0.12427      0.39276      0.32      0.7519
## Year2008     -0.38598      0.31068     -1.24      0.2151
## Year2009     -0.12174      0.37434     -0.33      0.7453
## Year2010     -0.45729      0.28968     -1.58      0.1156
## Year2011     -0.23930      0.28915     -0.83      0.4086
## Year2012     -0.31268      0.29869     -1.05      0.2961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.976
## Multiple R-squared:  0.0843, Adjusted R-squared:  0.0287
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 274 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.284  0.894   0.954   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      3.36e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 298"
## [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
## 168 125 157 124 178 172 165 132 168 180 228 233 259 284 278
## 2011 2012
## 343 376
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 141 97 128 103 131 101 135 109 139 142 199 197 229 229 231
## 2011 2012
## 298 312
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 138 92 122 99 122 94 130 99 129 123 188 184 218 218 215
## 2011 2012
## 279 288
## [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
```

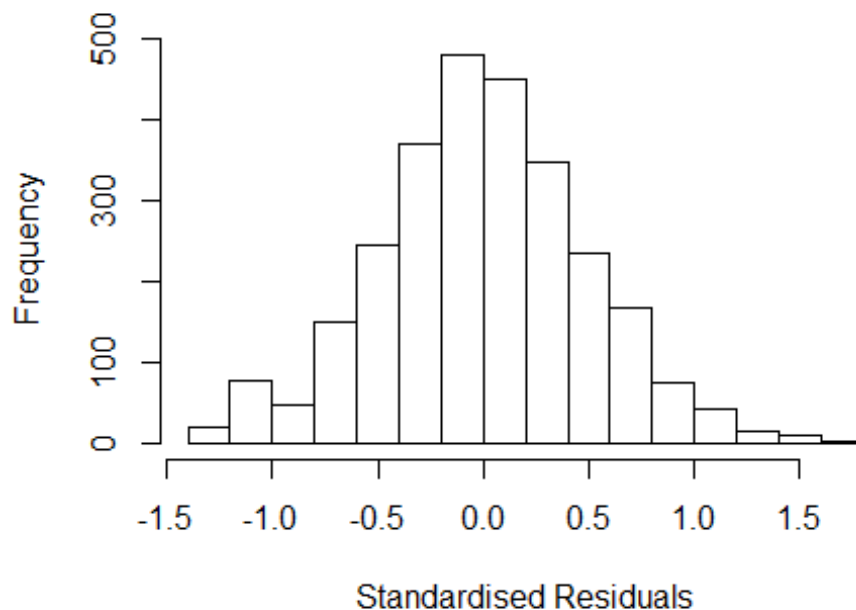


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.09630391557667"
## [1] "Male first author team size 2018 geometric mean: 1.88403942290815"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.00222503334538"
## [1] "Male last author team size 2018 geometric mean: 2.01311393383651"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18000, 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 1.246 1          1.116
## LastAuthorFemale  1.188 1          1.090
## UniqueAuthors    1.151 4          1.018
## Year              1.182 16         1.005
```


Residuals from first and last author and team size



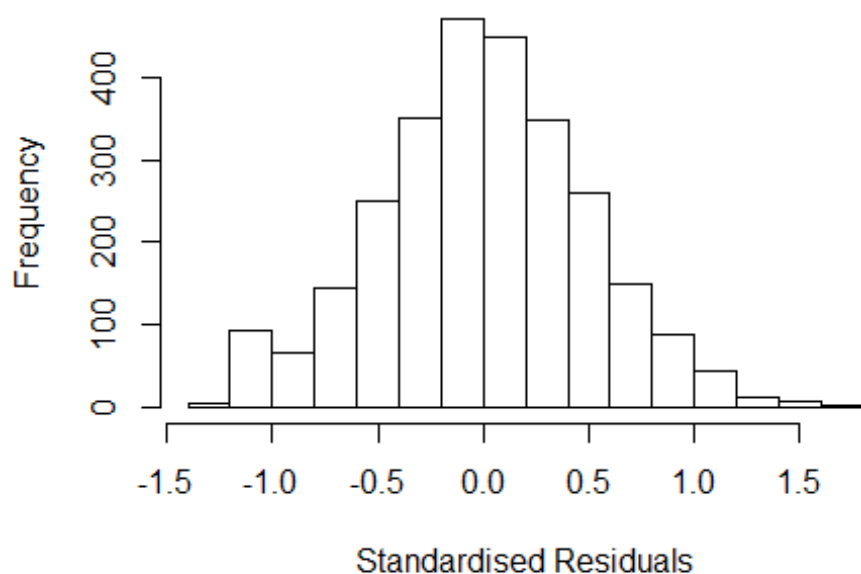
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29259 -0.30667 -0.00734  0.32053  1.71022
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.89178    0.05844   15.26 < 2e-16 ***
## FirstAuthorFemale1 -0.00557    0.02099   -0.27  0.7907
## LastAuthorFemale1  0.00801    0.02048    0.39  0.6958
## UniqueAuthors2     0.19092    0.02404    7.94 2.9e-15 ***
## UniqueAuthors3     0.19349    0.02784    6.95 4.6e-12 ***
## UniqueAuthors4     0.18788    0.03262    5.76 9.4e-09 ***
## UniqueAuthors5     0.25688    0.03926    6.54 7.2e-11 ***
## Year1997          0.03805    0.08157    0.47  0.6409
## Year1998          0.14043    0.07559    1.86  0.0633 .
## Year1999          0.19636    0.07892    2.49  0.0129 *
```

```

## Year2000      0.14759    0.07298    2.02    0.0432 *
## Year2001      0.11450    0.07112    1.61    0.1075
## Year2002      0.12738    0.07015    1.82    0.0695 .
## Year2003      0.14467    0.07552    1.92    0.0555 .
## Year2004      0.11763    0.06742    1.74    0.0812 .
## Year2005      0.13892    0.06978    1.99    0.0466 *
## Year2006      0.10222    0.06592    1.55    0.1211
## Year2007      0.14393    0.06600    2.18    0.0293 *
## Year2008      0.15516    0.06713    2.31    0.0209 *
## Year2009      0.15590    0.06661    2.34    0.0193 *
## Year2010      0.19437    0.06779    2.87    0.0042 **
## Year2011      0.15334    0.06475    2.37    0.0179 *
## Year2012      0.13561    0.06793    2.00    0.0460 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0494, Adjusted R-squared:  0.0417
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 236 weights are ~= 1. The remaining 2502 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.156  0.861  0.951   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      3.65e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.213 1      1.101
## LastAuthorFemale  1.174 1      1.084
## Year              1.072 16      1.002

```

Residuals from first and last author



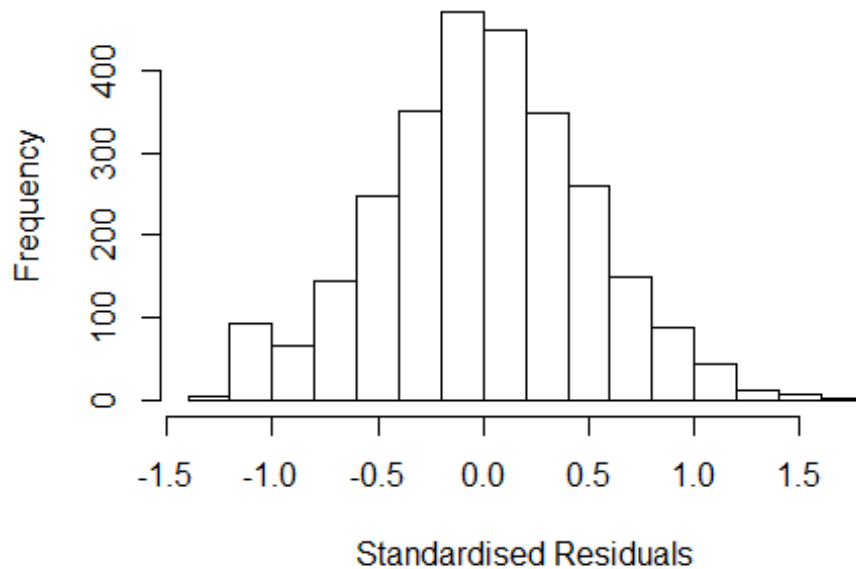
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22014 -0.32628 -0.00345  0.32402  1.65020
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.958532   0.059752   16.04 < 2e-16 ***
## FirstAuthorFemale1 0.021662   0.021129    1.03  0.30534
## LastAuthorFemale1 -0.000218   0.020860   -0.01  0.99166
## Year1997         0.082880   0.084907    0.98  0.32909
## Year1998         0.163270   0.078664    2.08  0.03803 *
## Year1999         0.217659   0.081079    2.68  0.00731 **
## Year2000         0.184437   0.074522    2.47  0.01339 *
## Year2001         0.151816   0.073522    2.06  0.03903 *
## Year2002         0.173392   0.071117    2.44  0.01483 *
## Year2003         0.185953   0.077218    2.41  0.01610 *
## Year2004         0.154031   0.069231    2.22  0.02617 *
## Year2005         0.193452   0.071458    2.71  0.00683 **
```

```

## Year2006          0.148669    0.067093    2.22  0.02678 *
## Year2007          0.188574    0.067445    2.80  0.00521 **
## Year2008          0.185499    0.068870    2.69  0.00711 **
## Year2009          0.189753    0.068240    2.78  0.00546 **
## Year2010          0.239946    0.069022    3.48  0.00052 ***
## Year2011          0.185879    0.066382    2.80  0.00514 **
## Year2012          0.192244    0.068930    2.79  0.00532 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.0103, Adjusted R-squared:  0.00375
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 253 weights are ~= 1. The remaining 2485 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.216  0.867  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      3.65e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1      1.025
## Year              1.051 16      1.002

```

Residuals from first author



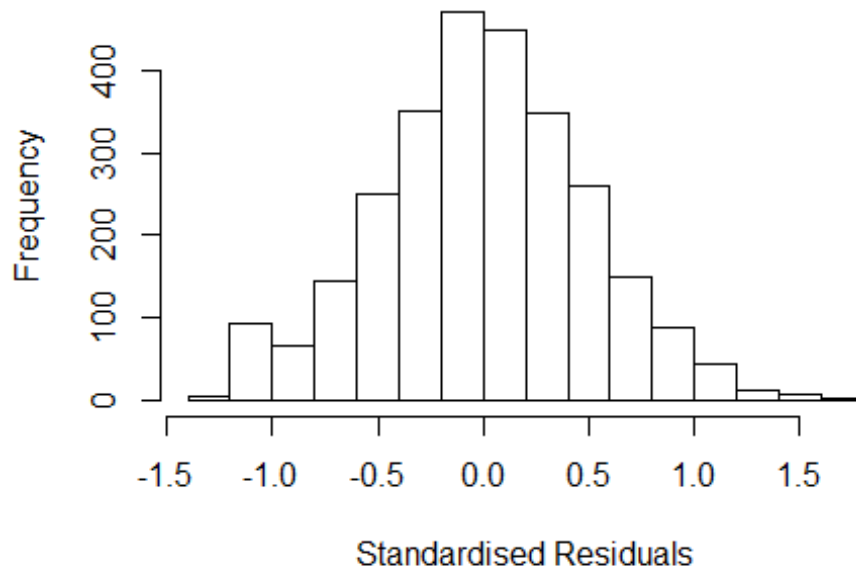
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21999 -0.32623 -0.00347 0.32407 1.65026
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9585 0.0594 16.13 < 2e-16 ***
## FirstAuthorFemale1 0.0216 0.0198 1.09 0.27596
## Year1997 0.0829 0.0849 0.98 0.32909
## Year1998 0.1633 0.0787 2.08 0.03804 *
## Year1999 0.2177 0.0811 2.68 0.00731 **
## Year2000 0.1845 0.0745 2.48 0.01337 *
## Year2001 0.1518 0.0735 2.06 0.03904 *
## Year2002 0.1734 0.0711 2.44 0.01479 *
## Year2003 0.1860 0.0771 2.41 0.01596 *
## Year2004 0.1541 0.0692 2.23 0.02614 *
## Year2005 0.1935 0.0715 2.71 0.00681 **
## Year2006 0.1487 0.0671 2.22 0.02679 *
```

```

## Year2007          0.1886      0.0675      2.80  0.00521 **
## Year2008          0.1855      0.0689      2.69  0.00712 **
## Year2009          0.1898      0.0682      2.78  0.00546 **
## Year2010          0.2400      0.0690      3.48  0.00052 ***
## Year2011          0.1859      0.0664      2.80  0.00513 **
## Year2012          0.1923      0.0689      2.79  0.00532 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.0103, Adjusted R-squared:  0.00412
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 253 weights are ~= 1. The remaining 2485 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.215  0.866  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      3.65e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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
## -1.21650 -0.32672 -0.00201 0.32755 1.64530
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96277 0.05964 16.14 <2e-16 ***
## LastAuthorFemale1 0.00992 0.01956 0.51 0.6121
## Year1997 0.08388 0.08505 0.99 0.3241
## Year1998 0.16393 0.07880 2.08 0.0376 *
## Year1999 0.21866 0.08098 2.70 0.0070 **
## Year2000 0.18582 0.07442 2.50 0.0126 *
## Year2001 0.15316 0.07352 2.08 0.0373 *
## Year2002 0.17399 0.07115 2.45 0.0145 *
## Year2003 0.19058 0.07675 2.48 0.0131 *
## Year2004 0.15563 0.06934 2.24 0.0249 *
## Year2005 0.19540 0.07136 2.74 0.0062 **
## Year2006 0.14959 0.06709 2.23 0.0258 *
```

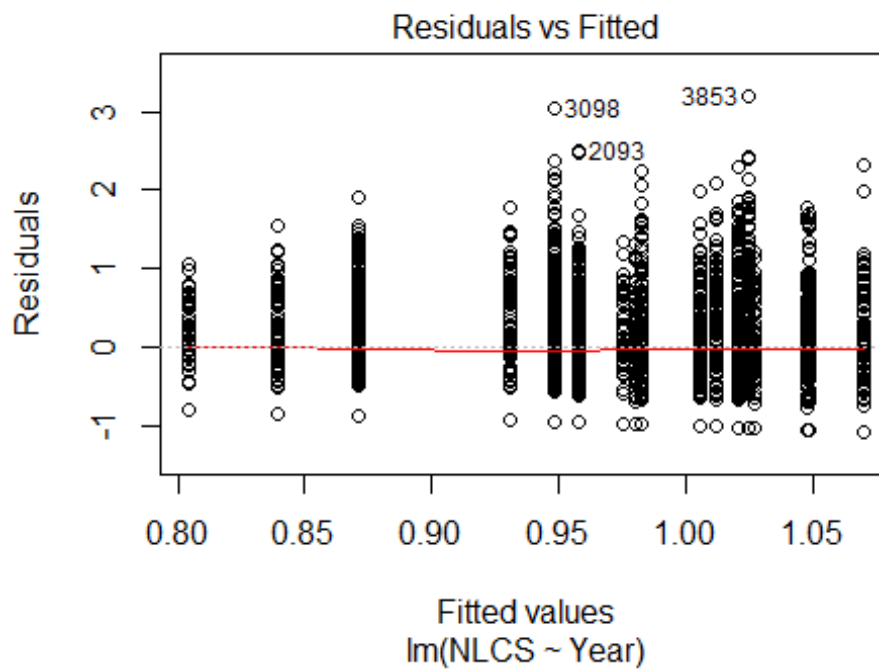
```

## Year2007      0.19009      0.06741      2.82      0.0048 **
## Year2008      0.18858      0.06869      2.75      0.0061 **
## Year2009      0.19340      0.06809      2.84      0.0045 **
## Year2010      0.24381      0.06881      3.54      0.0004 ***
## Year2011      0.18883      0.06624      2.85      0.0044 **
## Year2012      0.19496      0.06887      2.83      0.0047 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.00996,    Adjusted R-squared:  0.00377
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 257 weights are ~= 1. The remaining 2481 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.218  0.867  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      3.65e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2738"
## [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
## 103 98 109 105 146 118 165 155 142 190 217 196 351 366 411
## 2011 2012
## 416 442
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 79 83 87 110 90 139 134 120 157 185 175 306 315 365
## 2011 2012

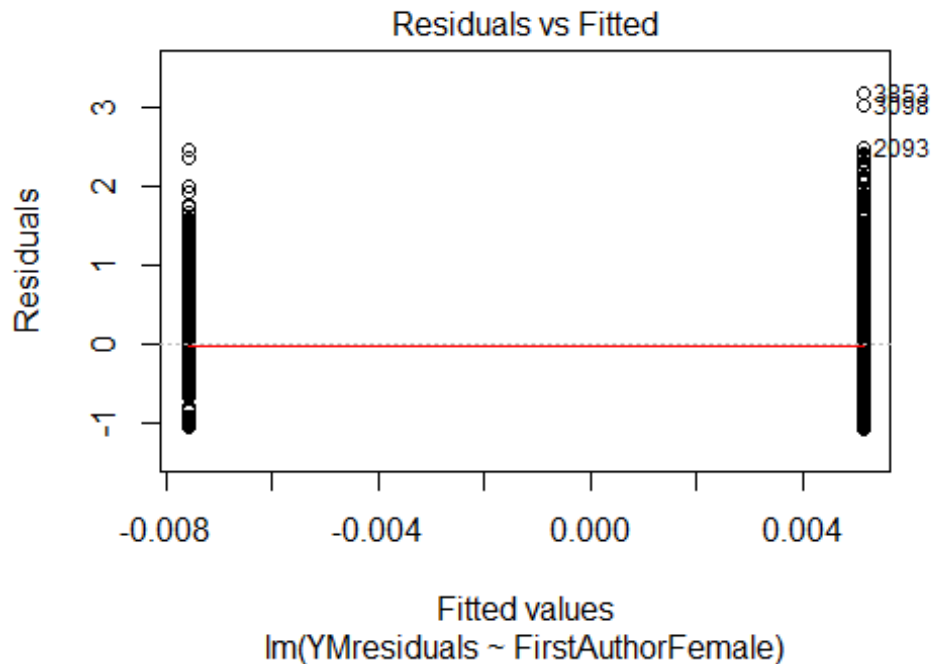
```



```
## 374 390
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 80 78 82 86 106 89 136 133 118 154 181 174 300 309 359
## 2011 2012
## 364 382
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 79, df = 16, p-value = 2e-10
```

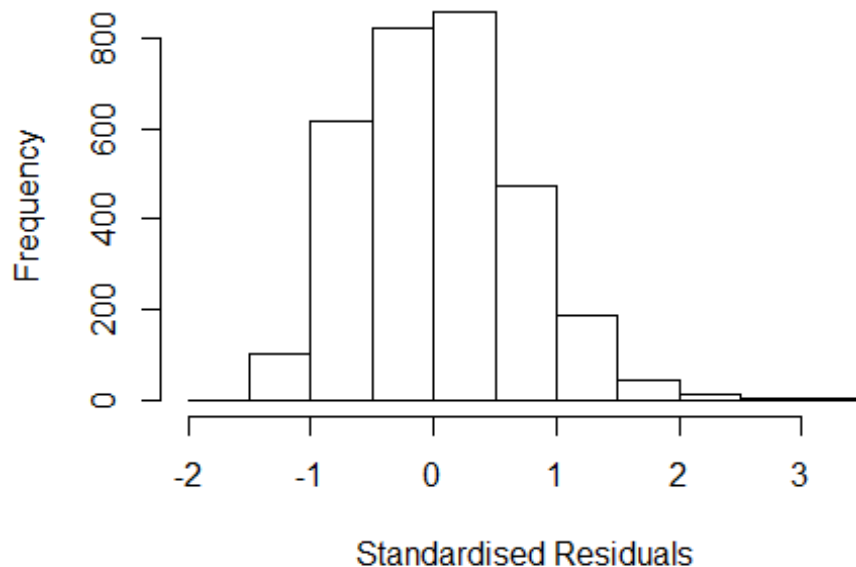


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.8, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 1.26778190245927"
## [1] "Male first author team size 2018 geometric mean: 1.14272149833952"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.18928743045063"
## [1] "Male last author team size 2018 geometric mean: 1.21660178632892"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, 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 5.931 1          2.435
## LastAuthorFemale  5.936 1          2.436
## UniqueAuthors    1.189 4          1.022
## Year             1.137 16          1.004
```

Residuals from first and last author and team size



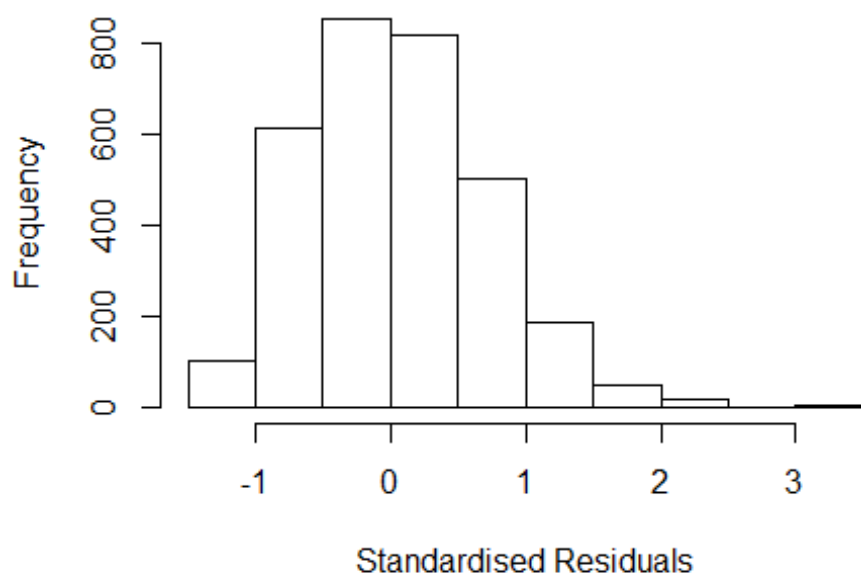
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2006 51749118540 3.422 2008    1202     4    2.502
## 2093 40449118934 3.462 2008    1202     4    2.546
## 3098 84555223668 3.991 2011    1202     2    3.144
## 3853 83655184689 4.215 2012    1202     3    3.281
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6051 -0.4565  0.0107  0.4582  3.2811
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95586    0.07354   13.00 < 2e-16 ***
## FirstAuthorFemale1 -0.04761    0.06149   -0.77  0.43885
## LastAuthorFemale1  0.05186    0.06192    0.84  0.40236
## UniqueAuthors2     0.19552    0.04431    4.41  1.1e-05 ***
## UniqueAuthors3     0.31718    0.10681    2.97  0.00300 **
## UniqueAuthors4     0.64924    0.17144    3.79  0.00016 ***
## UniqueAuthors5     0.12337    0.30140    0.41  0.68234
```

```

## Year1997          0.06648      0.10389      0.64  0.52229
## Year1998          -0.18543      0.09710     -1.91  0.05626 .
## Year1999          -0.00203      0.09441     -0.02  0.98281
## Year2000          -0.16052      0.09306     -1.72  0.08464 .
## Year2001           0.05343      0.09157      0.58  0.55963
## Year2002           0.07127      0.09902      0.72  0.47171
## Year2003          -0.06991      0.09554     -0.73  0.46435
## Year2004           0.05558      0.09697      0.57  0.56653
## Year2005          -0.00268      0.09756     -0.03  0.97806
## Year2006           0.00836      0.08813      0.09  0.92441
## Year2007          -0.03955      0.09428     -0.42  0.67486
## Year2008          -0.03978      0.08255     -0.48  0.62990
## Year2009           0.01115      0.08321      0.13  0.89343
## Year2010          -0.14079      0.08124     -1.73  0.08319 .
## Year2011          -0.10869      0.08568     -1.27  0.20469
## Year2012          -0.02197      0.08488     -0.26  0.79576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.0175
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 277 weights are ~= 1. The remaining 2854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8610 0.9520 0.9130 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      3.19e-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 5.901 1          2.429
## LastAuthorFemale 5.819 1          2.412
## Year              1.065 16          1.002

```

Residuals from first and last author



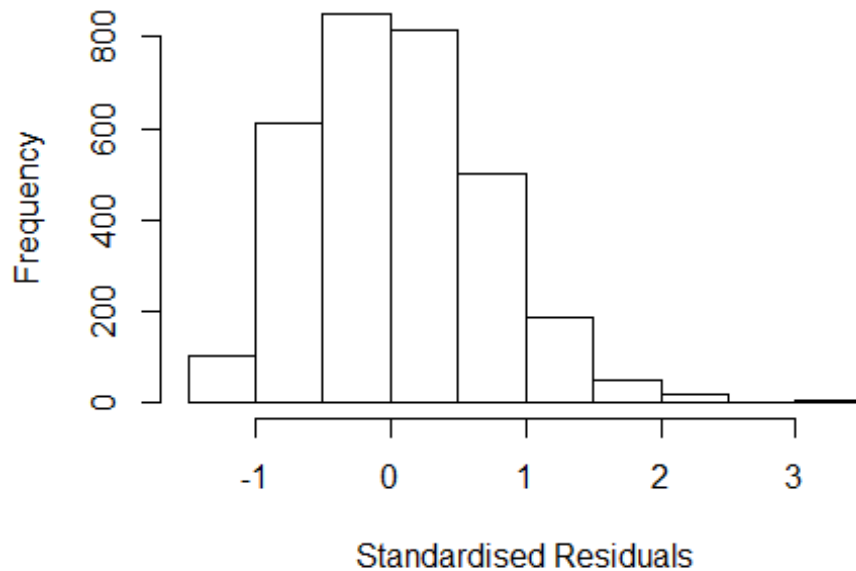
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2093 40449118934 3.462 2008    1202    4    2.520
## 3098 84555223668 3.991 2011    1202    2    3.105
## 3853 83655184689 4.215 2012    1202    3    3.241
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.06e+00 -4.55e-01 -5.40e-07  4.66e-01  3.24e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.980336   0.071325   13.74  <2e-16 ***
## FirstAuthorFemale1  0.019556   0.061648    0.32   0.751
## LastAuthorFemale1 -0.018245   0.061822   -0.30   0.768
## Year1997         0.052633   0.101748    0.52   0.605
## Year1998        -0.180231   0.095246   -1.89   0.059 .
## Year1999        -0.000201   0.091643    0.00   0.998
## Year2000        -0.153962   0.092266   -1.67   0.095 .
## Year2001         0.044256   0.090253    0.49   0.624
## Year2002         0.059004   0.097684    0.60   0.546
## Year2003        -0.073561   0.093814   -0.78   0.433
```

```

## Year2004          0.064562    0.095505    0.68    0.499
## Year2005          0.000546    0.095954    0.01    0.995
## Year2006          0.000207    0.086822    0.00    0.998
## Year2007         -0.036904    0.092721   -0.40    0.691
## Year2008         -0.038336    0.080637   -0.48    0.635
## Year2009          0.023627    0.081542    0.29    0.772
## Year2010         -0.127669    0.079129   -1.61    0.107
## Year2011         -0.093902    0.083684   -1.12    0.262
## Year2012         -0.005971    0.082981   -0.07    0.943
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.716
## Multiple R-squared:  0.0089, Adjusted R-squared:  0.00317
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 280 weights are ~= 1. The remaining 2851 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0044 0.8650 0.9530 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      3.19e-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 1.046 1      1.023
## Year              1.046 16      1.001

```

Residuals from first author



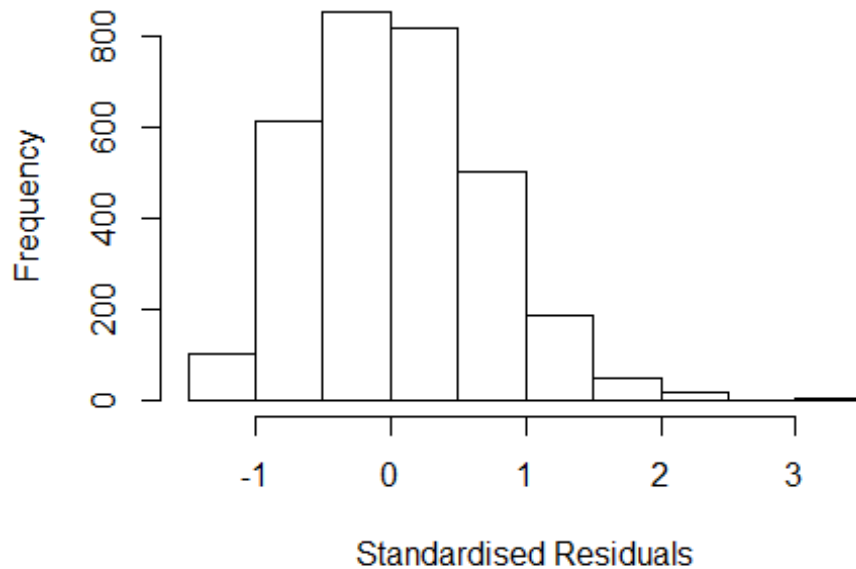
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2093 40449118934 3.462 2008    1202     4    2.520
## 3098 84555223668 3.991 2011    1202     2    3.105
## 3853 83655184689 4.215 2012    1202     3    3.241
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.047154 -0.457042  0.000561  0.469862  3.241323
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.80e-01   7.13e-02  13.74  <2e-16 ***
## FirstAuthorFemale1 3.21e-03   2.60e-02   0.12   0.902
## Year1997        5.30e-02   1.02e-01   0.52   0.602
## Year1998       -1.80e-01   9.52e-02  -1.89   0.059 .
## Year1999        1.36e-04   9.16e-02   0.00   0.999
## Year2000       -1.54e-01   9.22e-02  -1.67   0.094 .
## Year2001        4.43e-02   9.02e-02   0.49   0.624
## Year2002        5.90e-02   9.77e-02   0.60   0.546
## Year2003       -7.36e-02   9.38e-02  -0.78   0.433
## Year2004        6.41e-02   9.54e-02   0.67   0.501
```

```

## Year2005          2.05e-05  9.58e-02  0.00  1.000
## Year2006          4.34e-05  8.68e-02  0.00  1.000
## Year2007         -3.71e-02  9.27e-02 -0.40  0.689
## Year2008         -3.84e-02  8.06e-02 -0.48  0.634
## Year2009          2.43e-02  8.15e-02  0.30  0.766
## Year2010         -1.27e-01  7.91e-02 -1.61  0.107
## Year2011         -9.37e-02  8.36e-02 -1.12  0.263
## Year2012         -6.13e-03  8.30e-02 -0.07  0.941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.716
## Multiple R-squared:  0.00886,    Adjusted R-squared:  0.00345
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 278 weights are ~= 1. The remaining 2853 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0043 0.8640 0.9530 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      3.19e-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 1.031 1          1.016
## Year            1.031 16          1.001

```


Residuals from last author



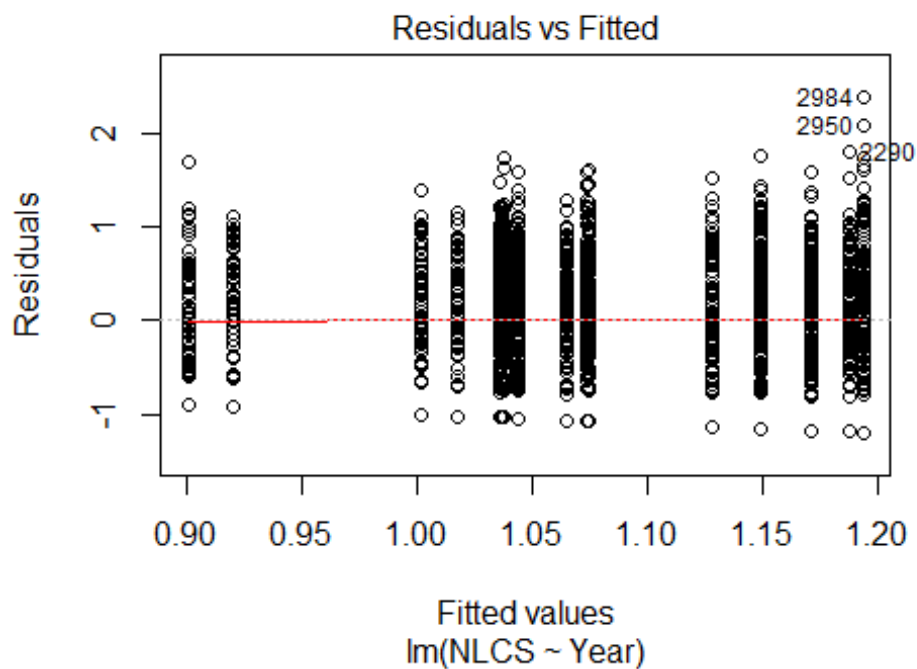
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2093 40449118934 3.462 2008    1202     4    2.520
## 3098 84555223668 3.991 2011    1202     2    3.105
## 3853 83655184689 4.215 2012    1202     3    3.241
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.045318 -0.458034  0.000327  0.469414  3.239676
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.980983   0.071240   13.77  <2e-16 ***
## LastAuthorFemale1 -0.000572   0.026033   -0.02   0.982
## Year1997        0.052691   0.101701    0.52   0.604
## Year1998       -0.180071   0.095236   -1.89   0.059 .
## Year1999        0.000262   0.091616    0.00   0.998
## Year2000       -0.154029   0.092218   -1.67   0.095 .
## Year2001        0.044423   0.090256    0.49   0.623
## Year2002        0.058963   0.097652    0.60   0.546
## Year2003       -0.073544   0.093791   -0.78   0.433
## Year2004        0.064335   0.095416    0.67   0.500
```

```

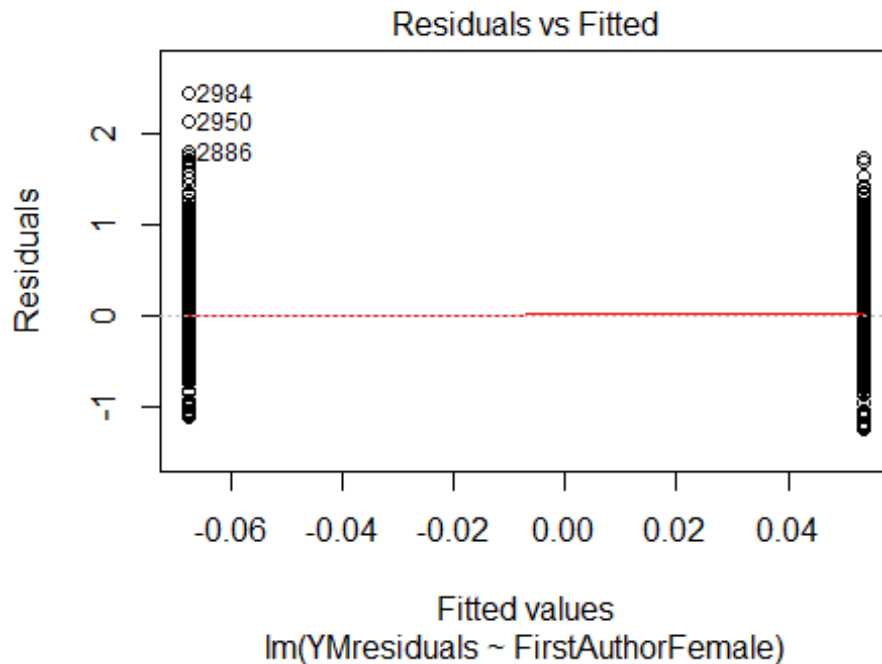
## Year2005      0.000156  0.095879  0.00  0.999
## Year2006      0.000217  0.086787  0.00  0.998
## Year2007     -0.036737  0.092692 -0.40  0.692
## Year2008     -0.037828  0.080564 -0.47  0.639
## Year2009      0.024775  0.081402  0.30  0.761
## Year2010     -0.126800  0.079061 -1.60  0.109
## Year2011     -0.093055  0.083522 -1.11  0.265
## Year2012     -0.005659  0.082926 -0.07  0.946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.715
## Multiple R-squared:  0.00886,    Adjusted R-squared:  0.00344
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 283 weights are ~= 1. The remaining 2848 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0043 0.8650 0.9530 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      3.19e-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: 3131"
## [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
## 102  90  96  99  102  117  138  136  158  143  168  161  223  228  197
## 2011 2012
## 300  264
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
##      89      74      81      86      79      94      115      118      126      122      136      140      181      185      160
## 2011 2012
## 259 223
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   84   71   75   83   78   92  110  115  124  114  130  136  172  179  151
## 2011 2012
## 251 212
## [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
```

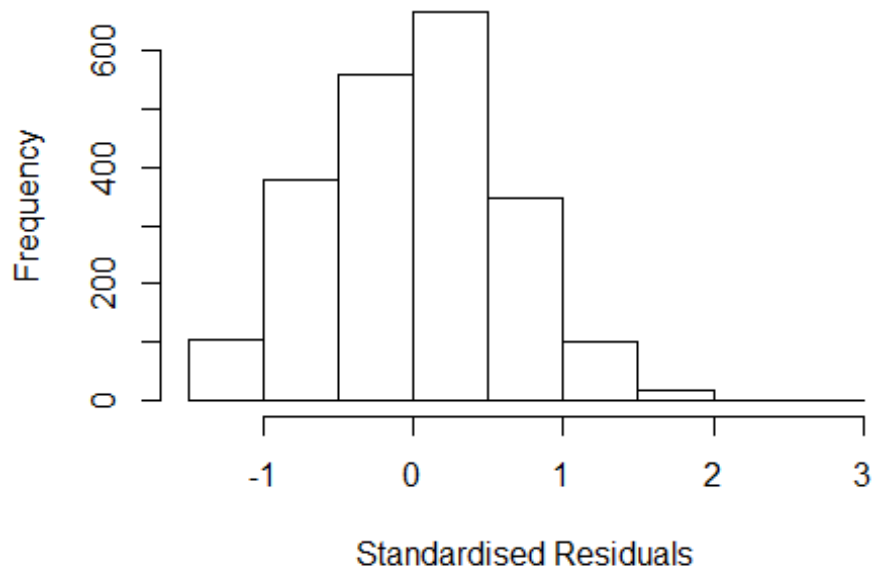


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.59, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.23860924726391"
## [1] "Male first author team size 2018 geometric mean: 1.55638237210383"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6600, p-value = 4e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.08371257011457"
## [1] "Male last author team size 2018 geometric mean: 1.79778707306387"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6000, 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 2.055 1          1.433
## LastAuthorFemale  1.865 1          1.365
## UniqueAuthors    1.492 4          1.051
## Year              1.294 16         1.008
```

Residuals from first and last author and team size



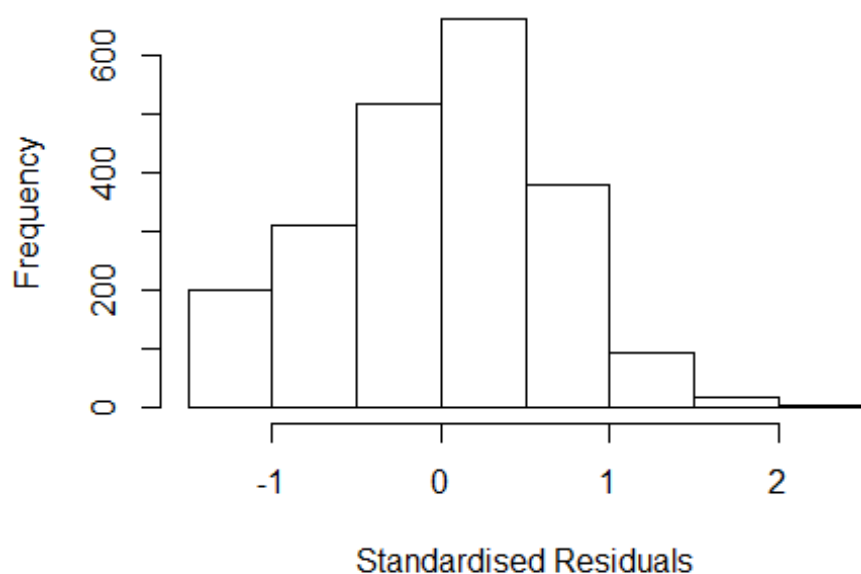
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2984 84859328881 3.575 2012      1203      3      2.591
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4726 -0.4352  0.0351  0.4301  2.5913
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8180    0.0617   13.25 < 2e-16 ***
## FirstAuthorFemale1 0.0363    0.0401    0.90  0.366
## LastAuthorFemale1 0.0213    0.0380    0.56  0.576
## UniqueAuthors2    0.2455    0.0384    6.39 2.1e-10 ***
## UniqueAuthors3    0.3825    0.0435    8.79 < 2e-16 ***
## UniqueAuthors4    0.4410    0.0528    8.36 < 2e-16 ***
## UniqueAuthors5    0.5183    0.0570    9.09 < 2e-16 ***
## Year1997          0.0866    0.0996    0.87  0.385
## Year1998         -0.0395    0.0934   -0.42  0.672
## Year1999          0.0694    0.0960    0.72  0.470
```

```

## Year2000      0.0561      0.0909      0.62      0.537
## Year2001      0.0988      0.0926      1.07      0.287
## Year2002      0.0582      0.0908      0.64      0.522
## Year2003      0.0804      0.0946      0.85      0.395
## Year2004      0.0685      0.0840      0.82      0.414
## Year2005      0.1058      0.0828      1.28      0.202
## Year2006      0.0620      0.0852      0.73      0.467
## Year2007      0.1757      0.0805      2.18      0.029 *
## Year2008      0.0550      0.0767      0.72      0.473
## Year2009      0.1560      0.0725      2.15      0.032 *
## Year2010      0.1528      0.0791      1.93      0.054 .
## Year2011      0.1320      0.0748      1.76      0.078 .
## Year2012      0.1657      0.0813      2.04      0.042 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.643
## Multiple R-squared:  0.0838, Adjusted R-squared:  0.0745
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 182 weights are ~= 1. The remaining 1995 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0678 0.8400 0.9510 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          4.59e-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 2.016 1          1.420
## LastAuthorFemale 1.981 1          1.407
## Year      1.090 16          1.003

```

Residuals from first and last author



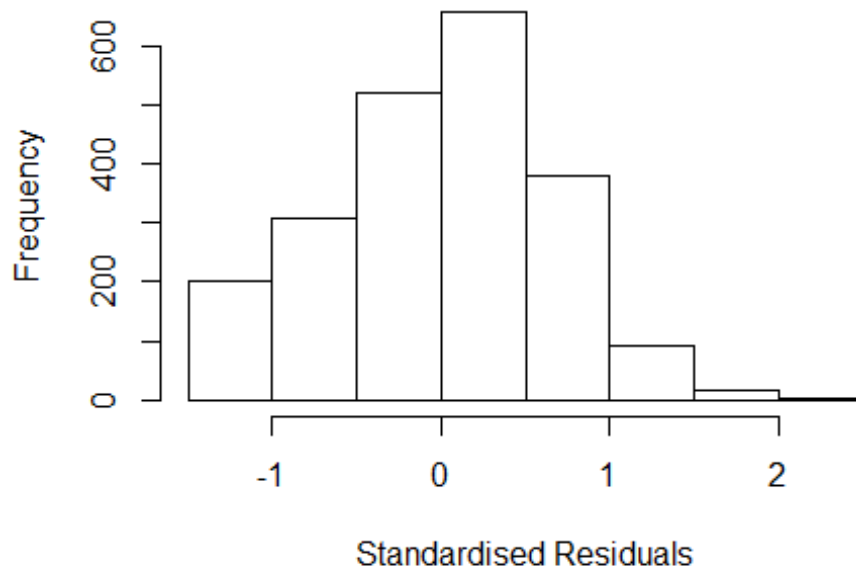
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2435 -0.4708 0.0508 0.4569 2.4758
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8505 0.0639 13.32 < 2e-16 ***
## FirstAuthorFemale1 0.1671 0.0420 3.98 7.2e-05 ***
## LastAuthorFemale1 -0.0434 0.0412 -1.05 0.2918
## Year1997 0.0996 0.1024 0.97 0.3305
## Year1998 -0.0231 0.0976 -0.24 0.8130
## Year1999 0.0761 0.0990 0.77 0.4423
## Year2000 0.0852 0.0983 0.87 0.3862
## Year2001 0.1246 0.0942 1.32 0.1862
## Year2002 0.1348 0.0934 1.44 0.1491
## Year2003 0.1377 0.0961 1.43 0.1521
## Year2004 0.1546 0.0854 1.81 0.0704 .
## Year2005 0.1339 0.0852 1.57 0.1164
```

```

## Year2006          0.1019      0.0880      1.16      0.2469
## Year2007          0.2066      0.0847      2.44      0.0148 *
## Year2008          0.1380      0.0795      1.74      0.0826 .
## Year2009          0.2259      0.0753      3.00      0.0027 **
## Year2010          0.2513      0.0808      3.11      0.0019 **
## Year2011          0.2104      0.0775      2.71      0.0067 **
## Year2012          0.2487      0.0843      2.95      0.0032 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.0233, Adjusted R-squared:  0.0151
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 2009 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.135  0.859   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      4.59e-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 1.055 1      1.027
## Year              1.055 16      1.002

```


Residuals from first author



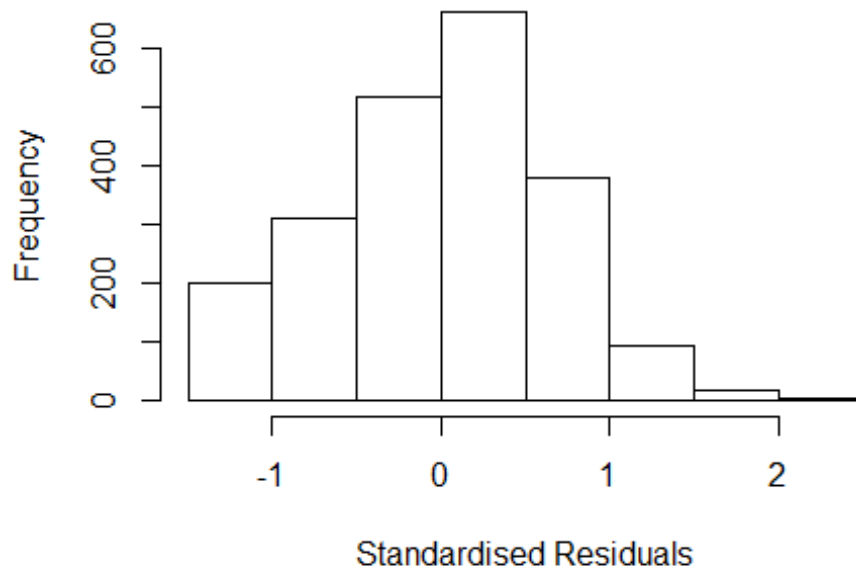
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2331 -0.4720 0.0463 0.4584 2.4775
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8436 0.0635 13.29 < 2e-16 ***
## FirstAuthorFemale1 0.1356 0.0304 4.46 8.5e-06 ***
## Year1997 0.1010 0.1024 0.99 0.3240
## Year1998 -0.0201 0.0972 -0.21 0.8361
## Year1999 0.0786 0.0993 0.79 0.4284
## Year2000 0.0847 0.0979 0.86 0.3872
## Year2001 0.1265 0.0944 1.34 0.1807
## Year2002 0.1403 0.0932 1.50 0.1325
## Year2003 0.1393 0.0961 1.45 0.1474
## Year2004 0.1565 0.0855 1.83 0.0673 .
## Year2005 0.1363 0.0853 1.60 0.1103
## Year2006 0.1039 0.0880 1.18 0.2379
```

```

## Year2007          0.2079      0.0847      2.45      0.0142 *
## Year2008          0.1397      0.0795      1.76      0.0790 .
## Year2009          0.2294      0.0751      3.05      0.0023 **
## Year2010          0.2532      0.0808      3.13      0.0017 **
## Year2011          0.2131      0.0773      2.76      0.0059 **
## Year2012          0.2539      0.0839      3.03      0.0025 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.0228, Adjusted R-squared:  0.0151
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 171 weights are ~= 1. The remaining 2006 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.133  0.860  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      4.59e-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 1.037 1      1.018
## Year              1.037 16      1.001

```

Residuals from last author



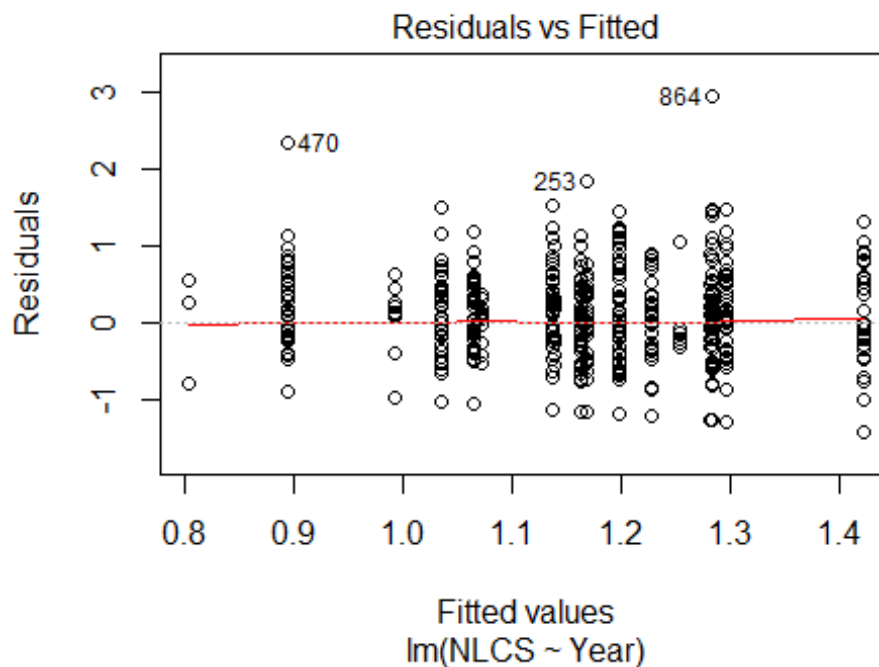
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2178 -0.4663 0.0401 0.4627 2.4329
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8757 0.0640 13.68 <2e-16 ***
## LastAuthorFemale1 0.0757 0.0300 2.53 0.0116 *
## Year1997 0.1014 0.1026 0.99 0.3234
## Year1998 -0.0244 0.0975 -0.25 0.8022
## Year1999 0.0784 0.1005 0.78 0.4354
## Year2000 0.0922 0.0973 0.95 0.3437
## Year2001 0.1321 0.0948 1.39 0.1637
## Year2002 0.1488 0.0941 1.58 0.1140
## Year2003 0.1441 0.0966 1.49 0.1359
## Year2004 0.1584 0.0857 1.85 0.0647 .
## Year2005 0.1383 0.0860 1.61 0.1081
## Year2006 0.1075 0.0882 1.22 0.2232
```

```

## Year2007          0.2072      0.0852      2.43      0.0151 *
## Year2008          0.1458      0.0804      1.81      0.0699 .
## Year2009          0.2429      0.0754      3.22      0.0013 **
## Year2010          0.2640      0.0812      3.25      0.0012 **
## Year2011          0.2159      0.0779      2.77      0.0056 **
## Year2012          0.2664      0.0845      3.15      0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.669
## Multiple R-squared:  0.0162, Adjusted R-squared:  0.00842
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 189 weights are ~= 1. The remaining 1988 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.158  0.858  0.946  0.911  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      4.59e-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: 2177"
## [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
##   11   13   16   15    8   28   39   55   41   54   55   56   44   59   50
## 2011 2012
##   68   72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6   10   10   11    3   12   32   41   35   47   49   51   34   50   41
## 2011 2012

```

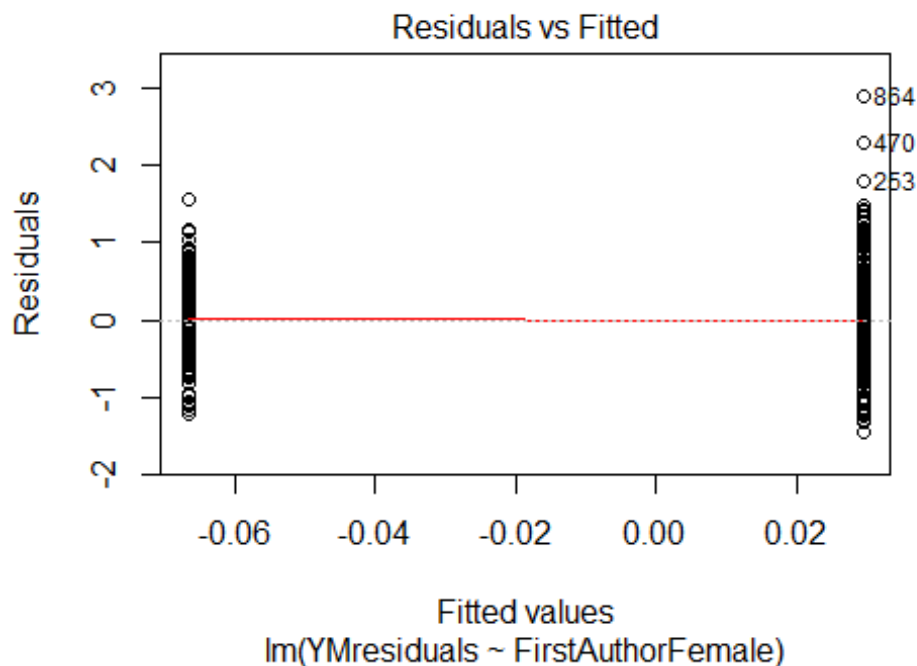
```
## 61 59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 9 9 11 3 11 31 39 33 44 46 49 34 48 40
## 2011 2012
## 58 57
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: 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 = 3.8, df = 1, p-value = 0.05
## [1] "Female first author team size 2018 geometric mean: 2.25541152420285"
## [1] "Male first author team size 2018 geometric mean: 2.15524645132765"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

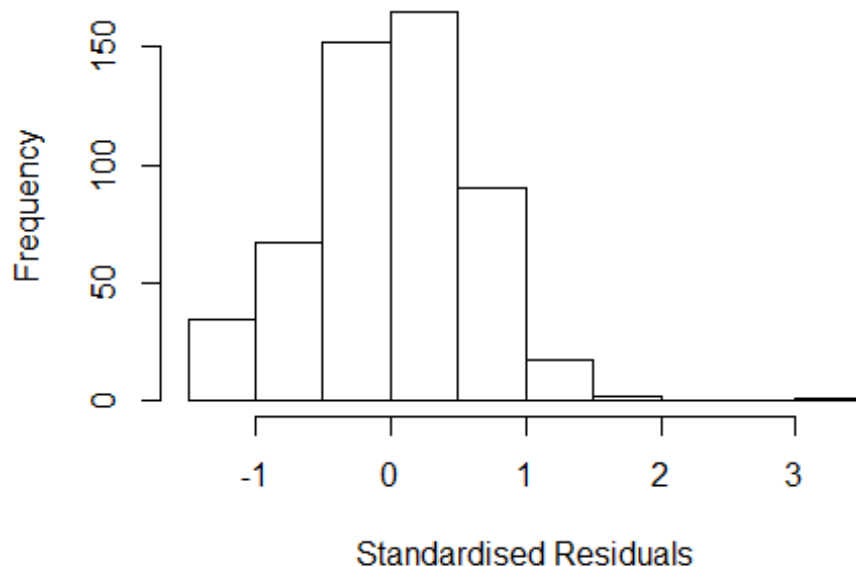
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 640, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.02904617817782"
## [1] "Male last author team size 2018 geometric mean: 2.38741876857339"

## 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 = 560, 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 1.858 1      1.363
## LastAuthorFemale  2.084 1      1.444
## UniqueAuthors    2.162 4      1.101
## Year              2.447 16      1.028
```

Residuals from first and last author and team size



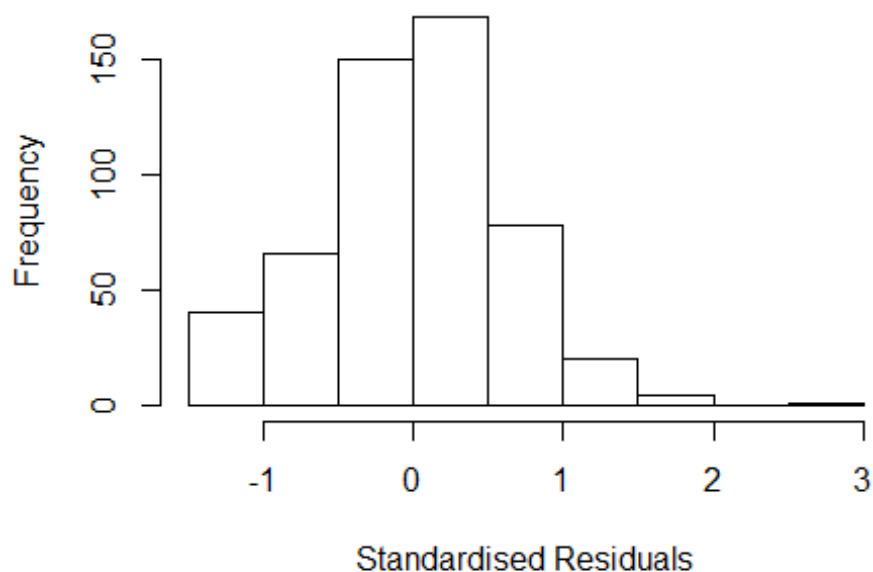
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 864 83655184689 4.215 2012    1202      3      3.118
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.388 -0.388  0.019  0.391  3.118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1174    0.1461   7.65 1.0e-13 ***
## FirstAuthorFemale1 -0.0486    0.0728  -0.67 0.50457
## LastAuthorFemale1 -0.0236    0.0804  -0.29 0.76942
## UniqueAuthors2     0.2914    0.0785   3.71 0.00023 ***
## UniqueAuthors3     0.4043    0.0805   5.02 7.1e-07 ***
## UniqueAuthors4     0.3891    0.1116   3.49 0.00053 ***
## UniqueAuthors5     0.2219    0.1181   1.88 0.06087 .
## Year1997          -0.2600    0.2020  -1.29 0.19869
## Year1998           0.0694    0.2058   0.34 0.73598
## Year1999          -0.1908    0.2033  -0.94 0.34827
```

```

## Year2000          -0.3813      0.4036    -0.94  0.34523
## Year2001          -0.1472      0.1695    -0.87  0.38555
## Year2002           0.2640      0.2145     1.23  0.21885
## Year2003          -0.0323      0.1775    -0.18  0.85572
## Year2004          -0.0104      0.1705    -0.06  0.95149
## Year2005          -0.0775      0.1807    -0.43  0.66823
## Year2006           0.0822      0.1734     0.47  0.63586
## Year2007          -0.3327      0.1718    -1.94  0.05338 .
## Year2008           0.0474      0.1760     0.27  0.78781
## Year2009          -0.2151      0.1706    -1.26  0.20797
## Year2010          -0.1631      0.1698    -0.96  0.33743
## Year2011          -0.1007      0.1727    -0.58  0.56018
## Year2012          -0.0207      0.1735    -0.12  0.90518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0758
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 515 is an outlier with |weight| = 0 ( < 0.00019);
## 42 weights are ~= 1. The remaining 485 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.384  0.860  0.954  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           1.89e-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 1.834 1 1.354
## LastAuthorFemale 1.921 1 1.386
## Year 1.379 16 1.010

```


Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 864 83655184689 4.215 2012    1202      3      2.953
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.491 -0.395  0.026  0.412  2.953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.234757   0.176752    6.99 8.9e-12 ***
## FirstAuthorFemale1  0.000941   0.075492    0.01  0.990
## LastAuthorFemale1 -0.081364   0.081463   -1.00  0.318
## Year1997         -0.272817   0.233636   -1.17  0.243
## Year1998          0.105841   0.250042    0.42  0.672
## Year1999         -0.091667   0.217726   -0.42  0.674
## Year2000         -0.396499   0.427597   -0.93  0.354
## Year2001         -0.146647   0.193767   -0.76  0.450
## Year2002          0.256096   0.232065    1.10  0.270
## Year2003         -0.081183   0.199715   -0.41  0.685
## Year2004         -0.050447   0.197380   -0.26  0.798
## Year2005         -0.050384   0.211684   -0.24  0.812
```

```

## Year2006          0.070841    0.203374    0.35    0.728
## Year2007          -0.382800    0.204001   -1.88    0.061 .
## Year2008          0.022625    0.201113    0.11    0.910
## Year2009          -0.164562    0.192349   -0.86    0.393
## Year2010          -0.141650    0.196915   -0.72    0.472
## Year2011          -0.065529    0.195419   -0.34    0.738
## Year2012          0.027268    0.196942    0.14    0.890
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.603
## Multiple R-squared:  0.056, Adjusted R-squared:  0.0226
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 515 is an outlier with |weight| = 0 ( < 0.00019);
## 41 weights are ~= 1. The remaining 486 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.341  0.867   0.954   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          1.89e-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 1.133  1          1.064
## Year              1.133 16          1.004

## [1] "List of 1 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 864 83655184689 4.215 2012      1202      3      2.953
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -1.4885 -0.4006  0.0253  0.4110  2.9674
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2125     0.1744   6.95 1.1e-11 ***
## FirstAuthorFemale1 -0.0464     0.0597  -0.78   0.437
## Year1997          -0.2403     0.2329  -1.03   0.303
## Year1998           0.1304     0.2481   0.53   0.600
## Year1999          -0.0693     0.2138  -0.32   0.746
## Year2000          -0.3542     0.4279  -0.83   0.408
## Year2001          -0.1386     0.1931  -0.72   0.473
## Year2002           0.2760     0.2303   1.20   0.231
## Year2003          -0.0634     0.1982  -0.32   0.749
## Year2004          -0.0401     0.1966  -0.20   0.838
## Year2005          -0.0492     0.2119  -0.23   0.817
## Year2006           0.0890     0.2013   0.44   0.659
## Year2007          -0.3606     0.2033  -1.77   0.077 .
## Year2008           0.0414     0.1988   0.21   0.835
## Year2009          -0.1499     0.1917  -0.78   0.435
## Year2010          -0.1178     0.1942  -0.61   0.544
## Year2011          -0.0580     0.1957  -0.30   0.767
## Year2012           0.0352     0.1966   0.18   0.858
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.603
## Multiple R-squared:  0.0544, Adjusted R-squared:  0.0229
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 515 is an outlier with |weight| = 0 ( < 0.00019);
## 43 weights are ~= 1. The remaining 484 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.335  0.865  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      1.89e-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"

## 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

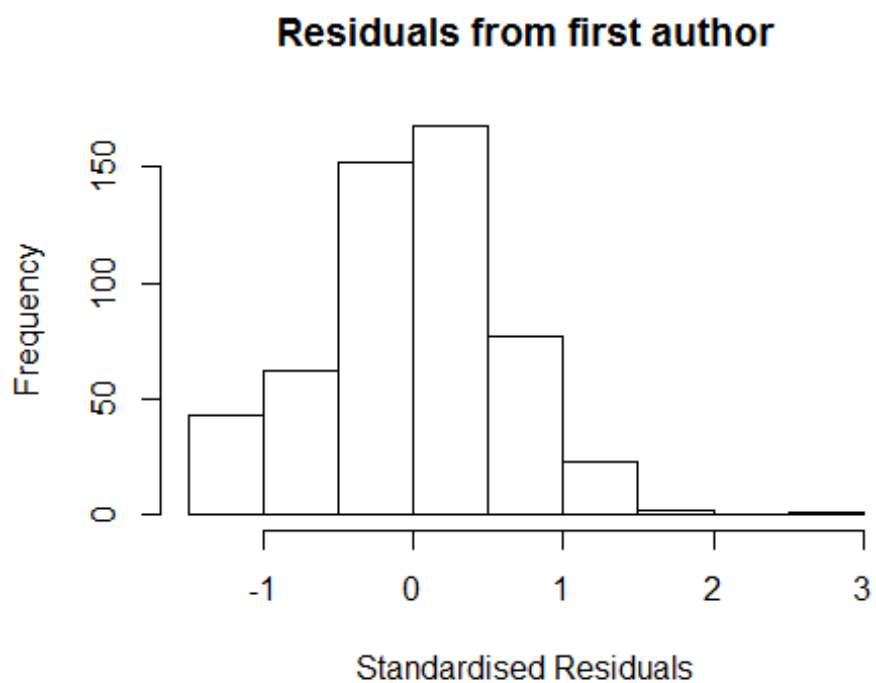
## [1] "Sample size for the above analysis: 528"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1205"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      5      3      4      1      4     10      6      7     11      9      8     12      9      4      6
## 2011 2012
##     14     24
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      2      2      1      1      2      5      3      4      7      7      5     11      8      3      5
## 2011 2012
##     12     19
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      2      2      1      1      2      5      3      4      7      7      5     11      8      3      5
## 2011 2012
##     11     19
## [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: 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 = 0.5, p-value = NA
## 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: 1"

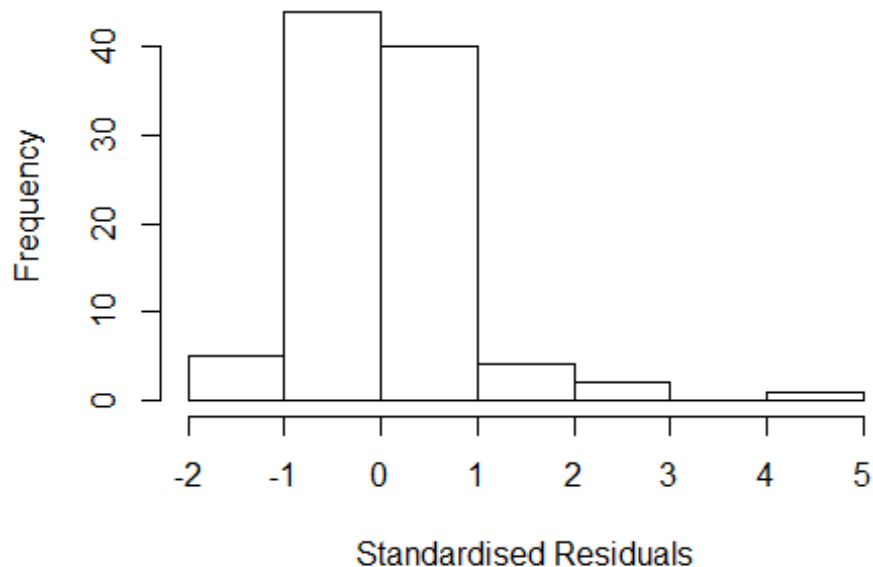
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0.5, p-value = NA
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.232e+01 1      3.51e+00
## LastAuthorFemale  3.026e+14 1      1.74e+07
## UniqueAuthors    -6.313e+14 1         NaN
## Year              -1.313e+29 16         NaN
```

Residuals from first and last author and team size



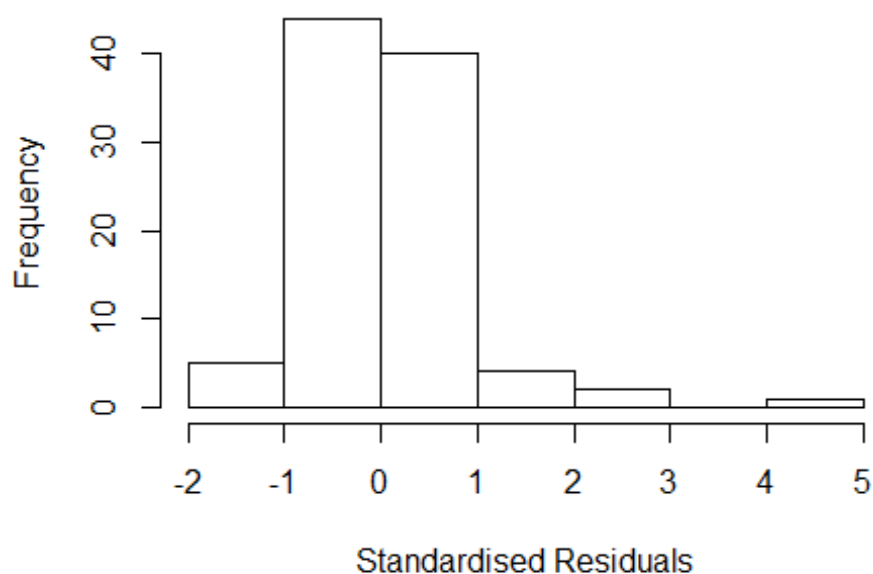
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 112 80052738768 4.134 2011      1205      1      2.620
## 131 84865530294 5.288 2012      1205      1      4.104
## 134 84871086462 3.772 2012      1200      2      2.835
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q       Max
## -1.51e+00 -3.36e-01  6.94e-16  3.92e-01  4.10e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.50670    0.38276   3.94  0.00018 ***
## FirstAuthorFemale1 -0.00386    0.34578  -0.01  0.99111
## LastAuthorFemale1 -0.24253    0.28629  -0.85  0.39957
## UniqueAuthors2   -0.24612    0.22135  -1.11  0.26967
## Year1997         -0.78120    0.38276  -2.04  0.04472 *
## Year1998         -0.44870    0.38276  -1.17  0.24474
## Year1999          0.62630    0.38276   1.64  0.10591
## Year2000        -1.11520    0.48691  -2.29  0.02477 *
```

```

## Year2001      -0.00857    0.41558   -0.02  0.98359
## Year2002      -0.15060    0.42047   -0.36  0.72122
## Year2003      -0.38466    0.54395   -0.71  0.48163
## Year2004      -0.78273    0.42923   -1.82  0.07215 .
## Year2005      -0.07366    0.44352   -0.17  0.86854
## Year2006      -0.62470    0.46686   -1.34  0.18486
## Year2007      -0.90402    0.39508   -2.29  0.02491 *
## Year2008      -0.46773    0.40356   -1.16  0.25009
## Year2009      -0.80811    0.80691   -1.00  0.31977
## Year2010      -0.67477    0.40855   -1.65  0.10274
## Year2011       0.00727    0.49624    0.01  0.98835
## Year2012      -0.32304    0.43456   -0.74  0.45954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.276, Adjusted R-squared:  0.0948
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(89,92) are outliers with |weight| = 0 ( < 0.001);
## 11 weights are ~= 1. The remaining 83 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0144 0.9010 0.9600 0.8980 0.9860 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          1.04e-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 4.886e+00 1          2.21
## LastAuthorFemale -6.440e+14 1          NaN
## Year -1.157e+15 16          NaN

```

Residuals from first and last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 112 80052738768 4.134 2011      1205      1      2.626
## 131 84865530294 5.288 2012      1205      1      4.111
## 134 84871086462 3.772 2012      1200      2      2.826
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q       Max
## -1.51e+00 -3.31e-01  7.77e-16  3.92e-01  4.11e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.49917    0.37676   3.98 0.00016 ***
## FirstAuthorFemale1  0.25107    0.23023   1.09 0.27888
## LastAuthorFemale1 -0.48241    0.24736  -1.95 0.05478 .
## Year1997        -0.77367    0.37676  -2.05 0.04342 *
## Year1998        -0.44117    0.37676  -1.17 0.24522
## Year1999         0.63383    0.37676   1.68 0.09656 .
## Year2000        -1.10767    0.48296  -2.29 0.02455 *
## Year2001        -0.06306    0.39464  -0.16 0.87347
## Year2002        -0.14788    0.41386  -0.36 0.72184
## Year2003        -0.37964    0.54578  -0.70 0.48878
```

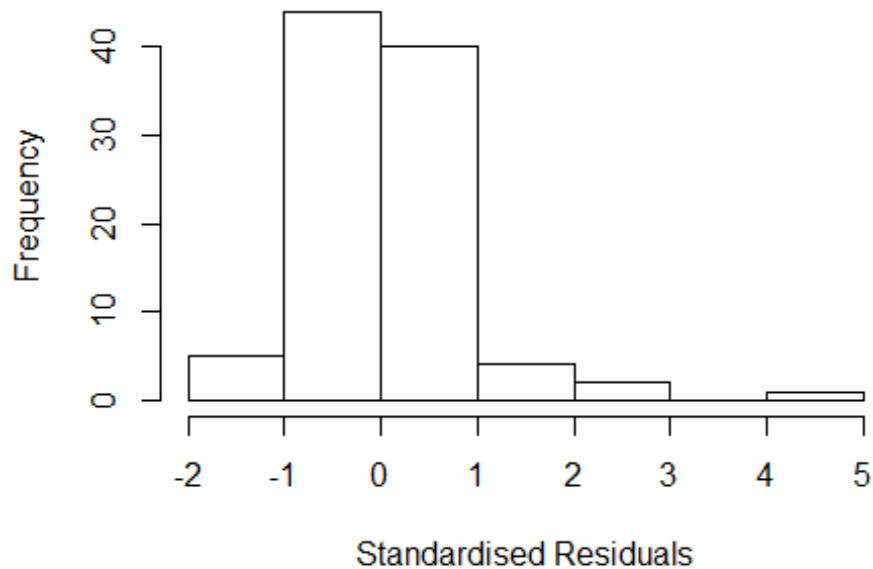


```

## Year2004      -0.77901    0.42640   -1.83  0.07158 .
## Year2005      -0.06430    0.43729   -0.15  0.88348
## Year2006      -0.61643    0.46257   -1.33  0.18658
## Year2007      -0.89747    0.38919   -2.31  0.02380 *
## Year2008      -0.46254    0.39850   -1.16  0.24934
## Year2009      -0.81396    0.81123   -1.00  0.31883
## Year2010      -0.67613    0.40439   -1.67  0.09858 .
## Year2011        0.00884    0.48889    0.02  0.98562
## Year2012      -0.32176    0.43021   -0.75  0.45680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.583
## Multiple R-squared:  0.277, Adjusted R-squared:  0.108
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(89,92) are outliers with |weight| = 0 ( < 0.001);
## 11 weights are ~= 1. The remaining 83 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0059 0.8920 0.9590 0.8940 0.9860 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      1.04e-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 2.217 1          1.489
## Year              2.217 16          1.025

```

Residuals from first author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 112 80052738768 4.134 2011      1205      1      2.626
## 131 84865530294 5.288 2012      1205      1      4.111
## 134 84871086462 3.772 2012      1200      2      2.826
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q       Max
## -1.51e+00 -3.55e-01 -3.89e-16  3.93e-01  4.16e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4913    0.3681   4.05 0.00012 ***
## FirstAuthorFemale1 -0.2156    0.1659  -1.30 0.19771
## Year1997          -0.7658    0.3681  -2.08 0.04077 *
## Year1998          -0.4333    0.3681  -1.18 0.24273
## Year1999           0.6417    0.3681   1.74 0.08522 .
## Year2000          -1.0998    0.4759  -2.31 0.02348 *
## Year2001          -0.0641    0.3857  -0.17 0.86835
## Year2002          -0.1451    0.4041  -0.36 0.72052
## Year2003          -0.3742    0.5422  -0.69 0.49222
## Year2004          -0.7758    0.4202  -1.85 0.06860 .
```

```

## Year2005          -0.0607      0.4292   -0.14  0.88789
## Year2006          -0.6089      0.4553   -1.34  0.18503
## Year2007          -0.8910      0.3806   -2.34  0.02176 *
## Year2008          -0.4582      0.3906   -1.17  0.24438
## Year2009          -0.8006      0.8037   -1.00  0.32224
## Year2010          -0.6778      0.3972   -1.71  0.09194 .
## Year2011           0.0154      0.4843     0.03  0.97471
## Year2012          -0.3646      0.4144   -0.88  0.38167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.269, Adjusted R-squared:  0.11
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(89,92) are outliers with |weight| = 0 ( < 0.001);
## 9 weights are ~ = 1. The remaining 85 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0084 0.8970 0.9580 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      1.04e-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 2.322 1          1.524
## Year            2.322 16          1.027
##
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 112 80052738768 4.134 2011      1205      1      2.626
## 131 84865530294 5.288 2012      1205      1      4.111
## 134 84871086462 3.772 2012      1200      2      2.826
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,

```

```

##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q        Max
## -1.51e+00 -3.18e-01 -5.83e-16  3.92e-01  4.13e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5058      0.3817   3.95  0.00017 ***
## LastAuthorFemale1 -0.2446      0.1648  -1.48  0.14178
## Year1997          -0.7803      0.3817  -2.04  0.04430 *
## Year1998          -0.4478      0.3817  -1.17  0.24429
## Year1999           0.6272      0.3817   1.64  0.10434
## Year2000          -1.1143      0.4863  -2.29  0.02465 *
## Year2001          -0.0616      0.4003  -0.15  0.87810
## Year2002          -0.1503      0.4200  -0.36  0.72150
## Year2003          -0.3841      0.5444  -0.71  0.48258
## Year2004          -0.7822      0.4293  -1.82  0.07227 .
## Year2005          -0.0716      0.4425  -0.16  0.87197
## Year2006          -0.6235      0.4661  -1.34  0.18489
## Year2007          -0.9032      0.3942  -2.29  0.02465 *
## Year2008          -0.4669      0.4029  -1.16  0.24999
## Year2009          -0.8118      0.8088  -1.00  0.31867
## Year2010          -0.6749      0.4085  -1.65  0.10252
## Year2011           0.0065      0.4949   0.01  0.98955
## Year2012          -0.3462      0.4251  -0.81  0.41788
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.273, Adjusted R-squared:  0.115
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(89,92) are outliers with |weight| = 0 ( < 0.001);
## 10 weights are ~= 1. The remaining 84 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0111 0.8930 0.9600 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          1.04e-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: 96"
## [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]"
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    2    4    4    2    2    6    3   10   15   13   15   16   26   15    9
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    3    3    2    1    4    3    7   12    9   12   13   24   13    7
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    2    3    2    1    4    3    6   12    9   12   13   22   13    7
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.25992104989487"
## [1] "Male first author team size 2018 geometric mean: 1.57113958920376"

## 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 = 65, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.35566197765995"
## [1] "Male last author team size 2018 geometric mean: 1.42180524661482"

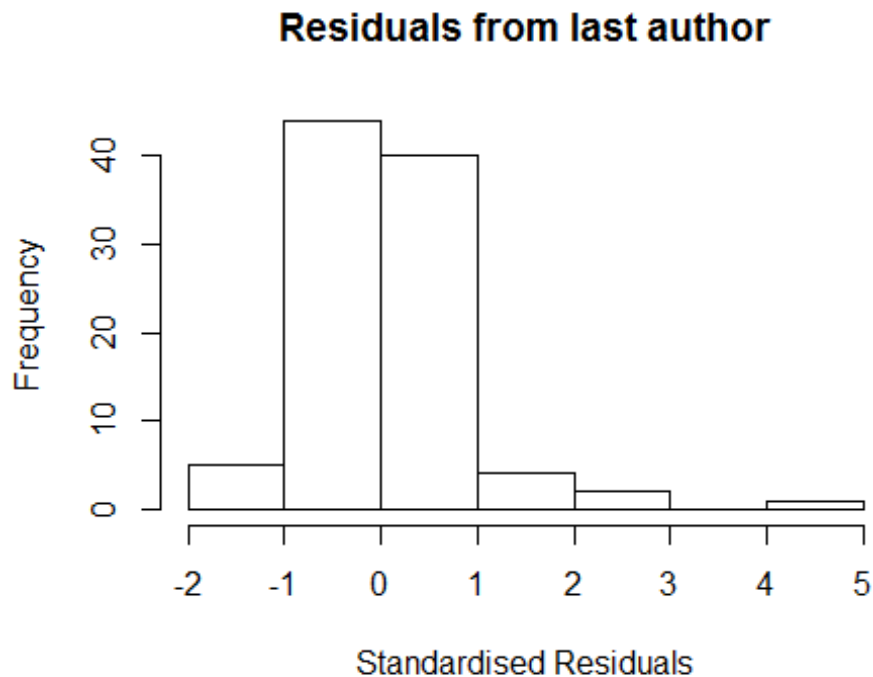
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 76, 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 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))
## FirstAuthorFemale NaN 1          NaN
## LastAuthorFemale  NaN 1          NaN
## UniqueAuthors    NaN 4          NaN
## Year             NaN 14         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 +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2492 -0.4781 -0.0543  0.5399  2.0229
##
```

```

## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.20e+00   2.57e-08  4.67e+07 < 2e-16 ***
## FirstAuthorFemale1 -3.84e-01   3.55e-01 -1.08e+00  0.2816
## LastAuthorFemale1  2.85e-01   3.61e-01  7.90e-01  0.4329
## UniqueAuthors2     5.82e-02   2.77e-01  2.10e-01  0.8341
## UniqueAuthors3     1.02e+00   3.49e-01  2.93e+00  0.0043 **
## UniqueAuthors4    -1.32e-02   6.73e-01 -2.00e-02  0.9843
## UniqueAuthors5    -2.07e-01   2.23e-01 -9.30e-01  0.3550
## Year1999          5.45e-01   4.38e-01  1.25e+00  0.2162
## Year2000         -6.62e-01   1.45e-01 -4.57e+00  1.6e-05 ***
## Year2001         -6.54e-01   1.84e-01 -3.56e+00  0.0006 ***
## Year2002         -1.20e+00   0.00e+00    -Inf < 2e-16 ***
## Year2003         -6.94e-01   2.80e-01 -2.48e+00  0.0150 *
## Year2004          1.09e-01   1.79e-01  6.10e-01  0.5432
## Year2005         -2.89e-01   3.26e-01 -8.90e-01  0.3780
## Year2006         -1.24e-01   2.63e-01 -4.70e-01  0.6371
## Year2007         -4.17e-01   2.93e-01 -1.42e+00  0.1589
## Year2008          4.53e-01   2.23e-01  2.03e+00  0.0450 *
## Year2009         -8.10e-01   2.51e-01 -3.23e+00  0.0018 **
## Year2010         -4.67e-02   1.85e-01 -2.50e-01  0.8015
## Year2011          1.49e-01   2.92e-01  5.10e-01  0.6115
## Year2012          1.48e-02   9.15e-01  2.00e-02  0.9871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.721
## Multiple R-squared:  0.286, Adjusted R-squared:  0.125
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~ = 1. The remaining 100 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.411  0.887  0.955   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          9.09e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01             5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##           0        1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)

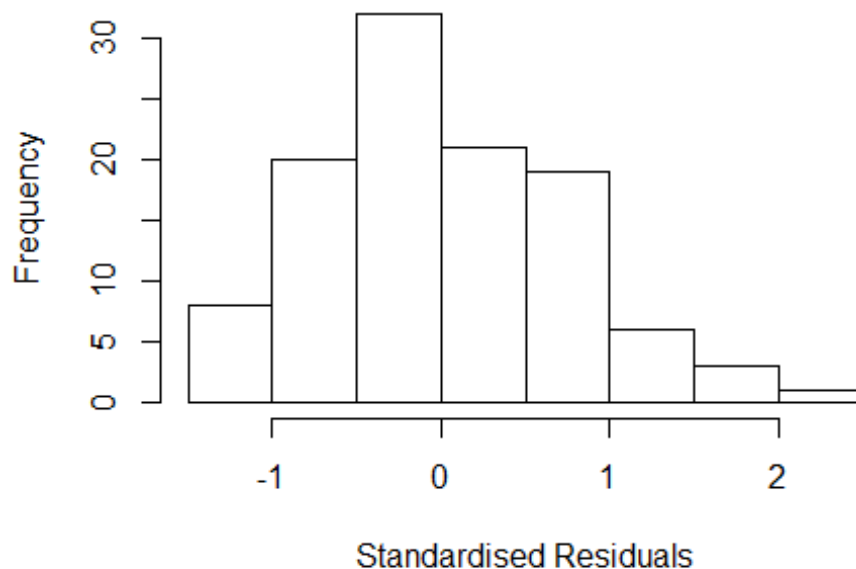
```

```
## [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 14         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
## -1.2638 -0.4420 -0.0718  0.5112  2.0469
##
```



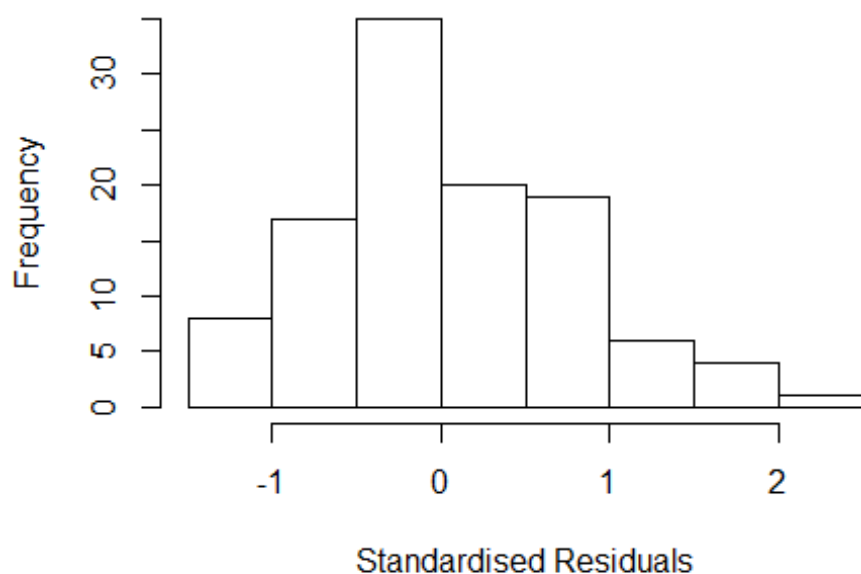
```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2000     0.0000      Inf < 2e-16 ***
## FirstAuthorFemale1 -0.3384     0.2390     -1.42  0.1602
## LastAuthorFemale1  0.2544     0.2375      1.07  0.2869
## Year1999          0.5455     0.4425      1.23  0.2207
## Year2000         -0.6636     0.1454     -4.56 1.5e-05 ***
## Year2001         -0.6325     0.1330     -4.76 7.2e-06 ***
## Year2002         -1.2000     0.0000     -Inf < 2e-16 ***
## Year2003         -0.6071     0.5401     -1.12  0.2639
## Year2004          0.4428     0.1852      2.39  0.0188 *
## Year2005         -0.2978     0.3261     -0.91  0.3634
## Year2006         -0.0263     0.2726     -0.10  0.9235
## Year2007         -0.4169     0.2794     -1.49  0.1391
## Year2008          0.4236     0.2053      2.06  0.0418 *
## Year2009         -0.8339     0.2473     -3.37  0.0011 **
## Year2010         -0.0402     0.1790     -0.22  0.8229
## Year2011          0.1478     0.2734      0.54  0.5901
## Year2012         -0.0243     0.9899     -0.02  0.9804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.694
## Multiple R-squared:  0.263, Adjusted R-squared:  0.136
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~ = 1. The remaining 101 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.364  0.877  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      9.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          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 6.914 1          2.630
## Year              6.914 14          1.071

## [1] "List of 0 outliers 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
## -1.2550 -0.4518 -0.0755  0.5090  2.0040
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.20e+00   0.00e+00      Inf < 2e-16 ***
## FirstAuthorFemale1 -1.41e-01   1.86e-01  -7.60e-01  0.45125
## Year1999         5.45e-01   4.43e-01   1.23e+00  0.22094
## Year2000        -7.09e-01   1.88e-01  -3.76e+00  0.00029 ***
## Year2001        -6.04e-01   1.16e-01  -5.22e+00  1.1e-06 ***
## Year2002       -1.20e+00   4.01e-08 -2.99e+07 < 2e-16 ***
## Year2003       -6.15e-01   5.51e-01  -1.12e+00  0.26698
## Year2004         4.62e-01   1.73e-01   2.67e+00  0.00904 **
```

```

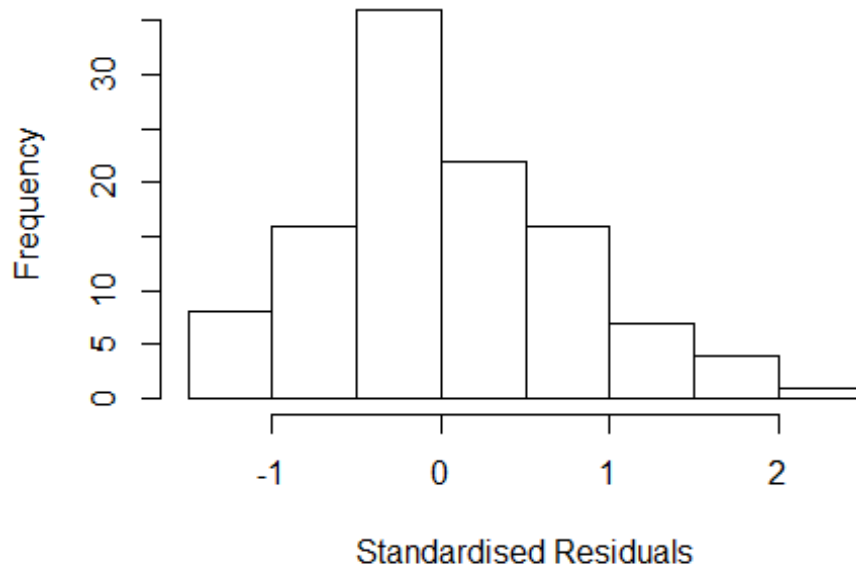
## Year2005      -2.96e-01   3.54e-01 -8.40e-01   0.40507
## Year2006      1.62e-02   2.69e-01  6.00e-02   0.95210
## Year2007     -3.89e-01   2.79e-01 -1.40e+00   0.16597
## Year2008      4.39e-01   2.05e-01  2.14e+00   0.03535 *
## Year2009     -7.91e-01   2.48e-01 -3.19e+00   0.00193 **
## Year2010     -4.42e-02   1.70e-01 -2.60e-01   0.79492
## Year2011      1.95e-01   2.87e-01  6.80e-01   0.49797
## Year2012     -1.79e-02   1.02e+00 -2.00e-02   0.98607
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.692
## Multiple R-squared:  0.255, Adjusted R-squared:  0.136
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~ = 1. The remaining 102 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.382  0.890   0.951   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      9.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          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

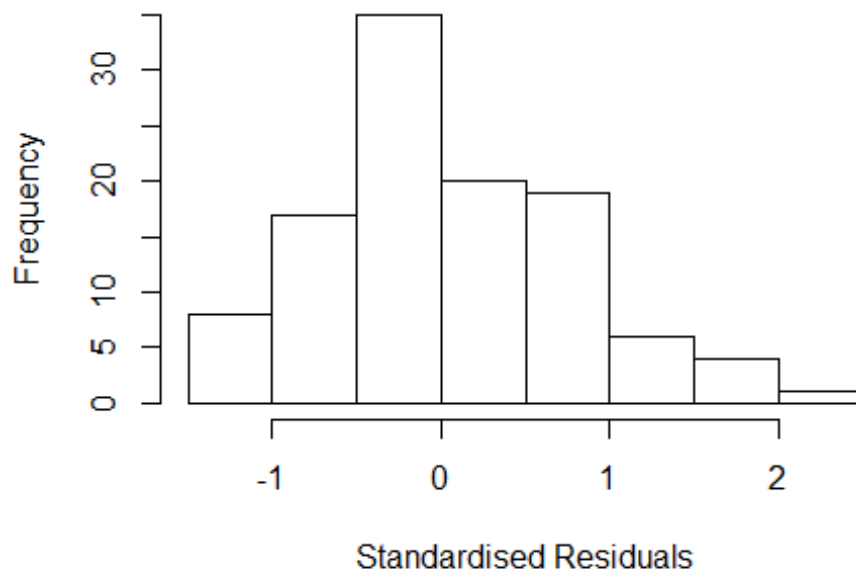
```

Residuals from first author



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale NaN 1          NaN
## Year           NaN 14          NaN
```

Residuals from last author

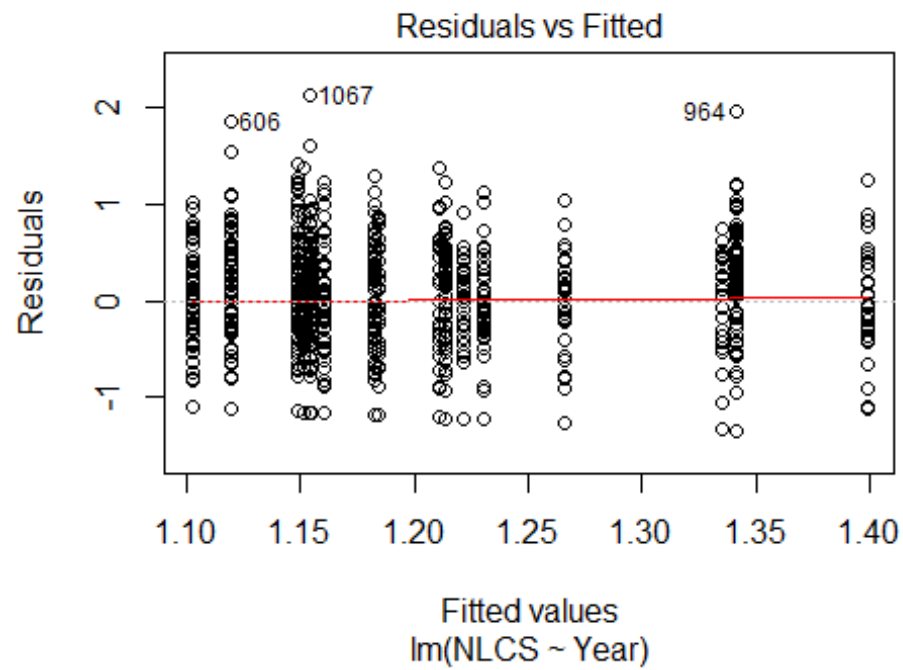


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3060 -0.4175 -0.0722 0.4878 2.1184
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20000 0.00000 Inf < 2e-16 ***
## LastAuthorFemale1 0.00259 0.18030 0.01 0.98859
## Year1999 0.54550 0.44435 1.23 0.22265
## Year2000 -0.80242 0.17779 -4.51 1.8e-05 ***
## Year2001 -0.67579 0.15065 -4.49 2.1e-05 ***
## Year2002 -1.20000 0.00000 -Inf < 2e-16 ***
## Year2003 -0.60655 0.52310 -1.16 0.24918
## Year2004 0.41426 0.20195 2.05 0.04302 *
## Year2005 -0.40795 0.34290 -1.19 0.23716
## Year2006 -0.08941 0.26086 -0.34 0.73255
## Year2007 -0.46310 0.27565 -1.68 0.09627 .
## Year2008 0.34203 0.20900 1.64 0.10508
## Year2009 -0.90545 0.22962 -3.94 0.00015 ***
## Year2010 -0.13657 0.14359 -0.95 0.34399
## Year2011 0.10339 0.30376 0.34 0.73435
## Year2012 -0.05440 0.98628 -0.06 0.95613
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.683
## Multiple R-squared: 0.252, Adjusted R-squared: 0.132
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~ 1. The remaining 102 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.316 0.880 0.949 0.900 0.983 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 9.09e-04 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
```

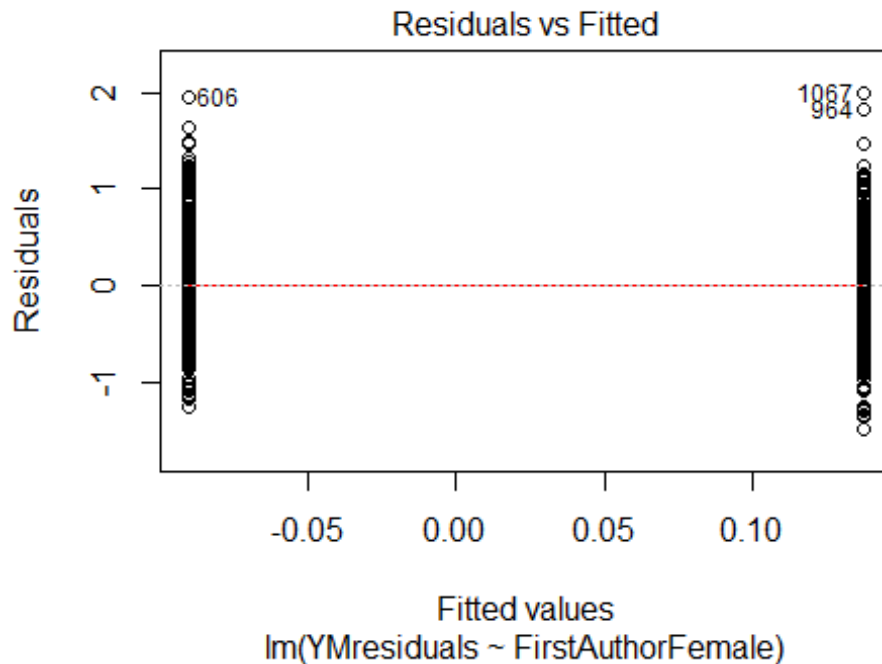
```

##           500           50           2           1           1000           200
## trace.lev      mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 110"
## [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
##   35   33   37   35   49   51   33   32   39   42   74   71   84   80  104
## 2011 2012
##   85  113
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   31   29   28   46   39   26   28   32   40   63   62   77   66   93
## 2011 2012
##   77   99
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   28   30   28   28   46   38   25   27   30   35   61   59   73   65   91
## 2011 2012
##   76   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 = 33, df = 16, p-value = 0.007

```

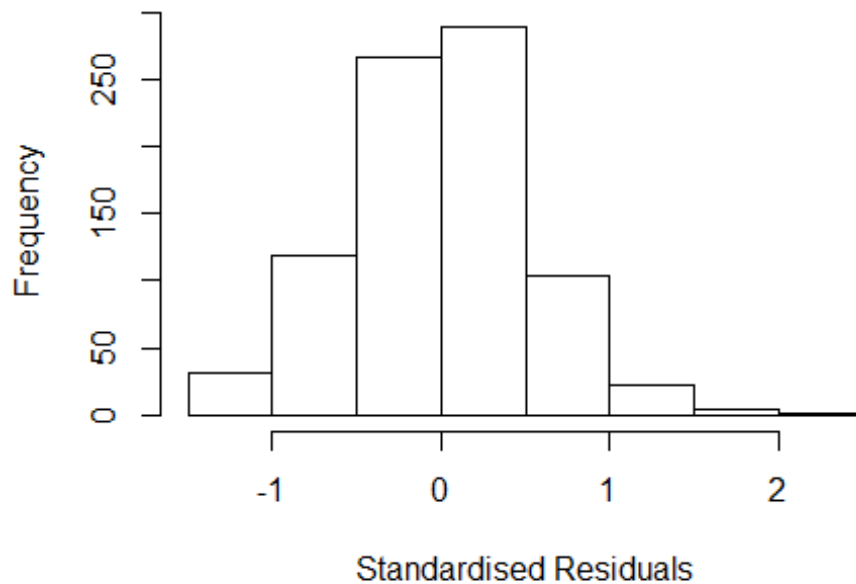


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.015, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 1.61555028881957"
## [1] "Male first author team size 2018 geometric mean: 1.21808969997177"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.45450017600234"
## [1] "Male last author team size 2018 geometric mean: 1.37595023737398"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, 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 1.833 1          1.354
## LastAuthorFemale  2.017 1          1.420
## UniqueAuthors    1.489 4          1.051
## Year             1.606 16          1.015
```


Residuals from first and last author and team size



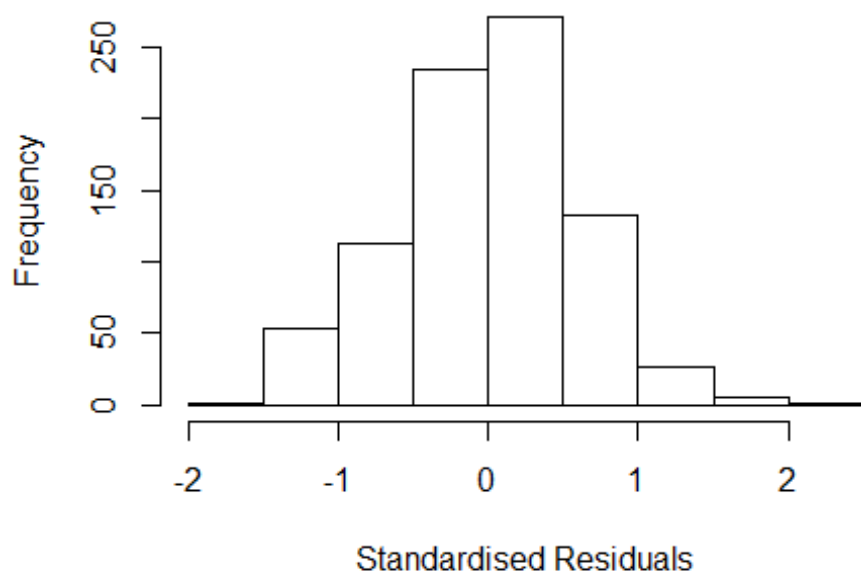
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.47598 -0.35190 0.00248 0.34614 2.22323
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9589 0.1172 8.18 1.1e-15 ***
## FirstAuthorFemale1 0.0849 0.0508 1.67 0.095 .
## LastAuthorFemale1 0.0589 0.0519 1.14 0.257
## UniqueAuthors2 0.3191 0.0553 5.77 1.1e-08 ***
## UniqueAuthors3 0.4705 0.0746 6.30 4.7e-10 ***
## UniqueAuthors4 0.6248 0.0667 9.36 < 2e-16 ***
## UniqueAuthors5 0.6146 0.0690 8.91 < 2e-16 ***
## Year1997 0.2673 0.1540 1.74 0.083 .
## Year1998 0.0794 0.1465 0.54 0.588
## Year1999 0.0709 0.1427 0.50 0.620
```

```

## Year2000          0.0546      0.1441      0.38      0.705
## Year2001         -0.0298      0.1400     -0.21      0.831
## Year2002          0.2439      0.1513      1.61      0.107
## Year2003          0.0998      0.1516      0.66      0.511
## Year2004         -0.0460      0.1471     -0.31      0.755
## Year2005          0.0824      0.1379      0.60      0.550
## Year2006         -0.0478      0.1274     -0.38      0.707
## Year2007          0.0403      0.1302      0.31      0.757
## Year2008         -0.0237      0.1381     -0.17      0.864
## Year2009          0.0380      0.1371      0.28      0.782
## Year2010         -0.0302      0.1305     -0.23      0.817
## Year2011          0.1390      0.1465      0.95      0.343
## Year2012         -0.0430      0.1328     -0.32      0.746
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.183, Adjusted R-squared:  0.161
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 758 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0408 0.8720 0.9490 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      1.19e-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 2.476 1      1.573
## LastAuthorFemale  2.519 1      1.587
## Year              1.144 16      1.004

```

Residuals from first and last author



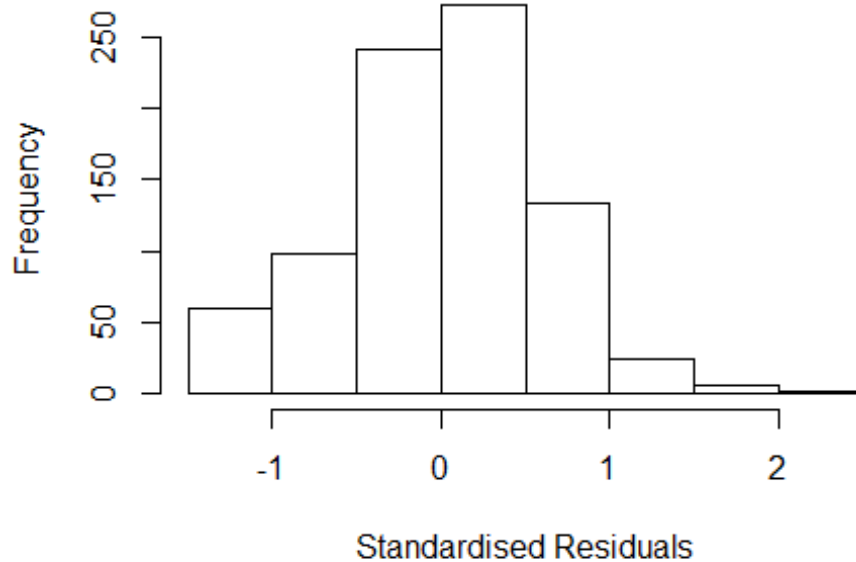
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5074 -0.3912 0.0268 0.3956 2.0340
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09841 0.13384 8.21 8.8e-16 ***
## FirstAuthorFemale1 0.15577 0.06531 2.38 0.017 *
## LastAuthorFemale1 0.09382 0.06566 1.43 0.153
## Year1997 0.24333 0.16757 1.45 0.147
## Year1998 0.06666 0.16252 0.41 0.682
## Year1999 0.00992 0.15799 0.06 0.950
## Year2000 -0.03855 0.16068 -0.24 0.810
## Year2001 -0.10226 0.16149 -0.63 0.527
## Year2002 0.18302 0.17152 1.07 0.286
## Year2003 0.05805 0.15579 0.37 0.710
## Year2004 -0.04918 0.16802 -0.29 0.770
## Year2005 0.00915 0.15675 0.06 0.953
```

```

## Year2006      -0.08363    0.14799   -0.57    0.572
## Year2007      0.00890    0.15036    0.06    0.953
## Year2008     -0.05829    0.15487   -0.38    0.707
## Year2009     -0.01763    0.15460   -0.11    0.909
## Year2010     -0.06909    0.15100   -0.46    0.647
## Year2011      0.15938    0.16071    0.99    0.322
## Year2012     -0.09902    0.15143   -0.65    0.513
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0567, Adjusted R-squared:  0.036
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 759 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.206  0.878  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      1.19e-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 1.07 1      1.035
## Year              1.07 16      1.002

```

Residuals from first author



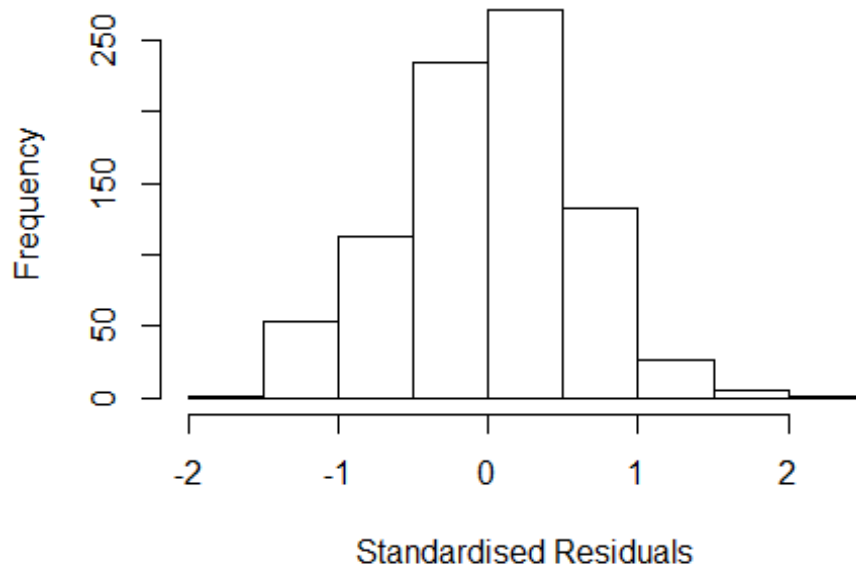
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4859 -0.3976 0.0242 0.3896 2.0446
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10872 0.13185 8.41 < 2e-16 ***
## FirstAuthorFemale1 0.22296 0.04286 5.20 2.5e-07 ***
## Year1997 0.23558 0.16520 1.43 0.15
## Year1998 0.06863 0.16224 0.42 0.67
## Year1999 -0.00184 0.15522 -0.01 0.99
## Year2000 -0.03508 0.15986 -0.22 0.83
## Year2001 -0.09597 0.16095 -0.60 0.55
## Year2002 0.18044 0.16987 1.06 0.29
## Year2003 0.05679 0.15575 0.36 0.72
## Year2004 -0.05240 0.16665 -0.31 0.75
## Year2005 0.00421 0.15571 0.03 0.98
## Year2006 -0.08283 0.14729 -0.56 0.57
```

```

## Year2007          0.00705    0.14846    0.05    0.96
## Year2008          -0.06519    0.15338   -0.42    0.67
## Year2009          -0.01856    0.15375   -0.12    0.90
## Year2010          -0.07075    0.14971   -0.47    0.64
## Year2011           0.15419    0.15883    0.97    0.33
## Year2012          -0.09327    0.15102   -0.62    0.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.59
## Multiple R-squared:  0.054, Adjusted R-squared:  0.0344
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.206  0.883   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      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 1.094 1          1.046
## Year              1.094 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
## -1.4899 -0.3951 0.0318 0.3857 2.0617
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10793 0.13610 8.14 1.5e-15 ***
## LastAuthorFemale1 0.20792 0.04302 4.83 1.6e-06 ***
## Year1997 0.25237 0.17107 1.48 0.14
## Year1998 0.06560 0.16438 0.40 0.69
## Year1999 0.01951 0.15918 0.12 0.90
## Year2000 -0.04147 0.16233 -0.26 0.80
## Year2001 -0.09278 0.16242 -0.57 0.57
## Year2002 0.19773 0.17206 1.15 0.25
## Year2003 0.06219 0.15647 0.40 0.69
## Year2004 -0.03902 0.17162 -0.23 0.82
## Year2005 0.01614 0.15797 0.10 0.92
## Year2006 -0.08484 0.14949 -0.57 0.57
```

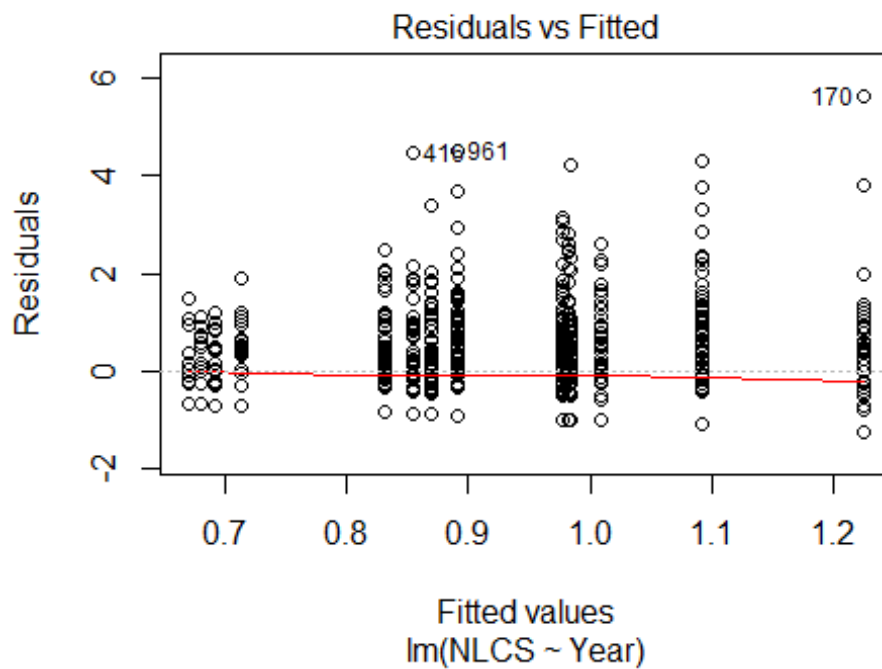
```

## Year2007      0.02139      0.15262      0.14      0.89
## Year2008     -0.04583      0.15654     -0.29      0.77
## Year2009     -0.00997      0.15535     -0.06      0.95
## Year2010     -0.05733      0.15266     -0.38      0.71
## Year2011      0.17403      0.16307      1.07      0.29
## Year2012     -0.09451      0.15204     -0.62      0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0495, Adjusted R-squared:  0.0297
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.191  0.878  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      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: 837"
## [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
##   21   25   16   25   31   42   65   82   67   71   78   97  120  102   93
## 2011 2012
##  156  177
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   22   10   21   23   34   54   65   56   61   65   87  102   86   84
## 2011 2012

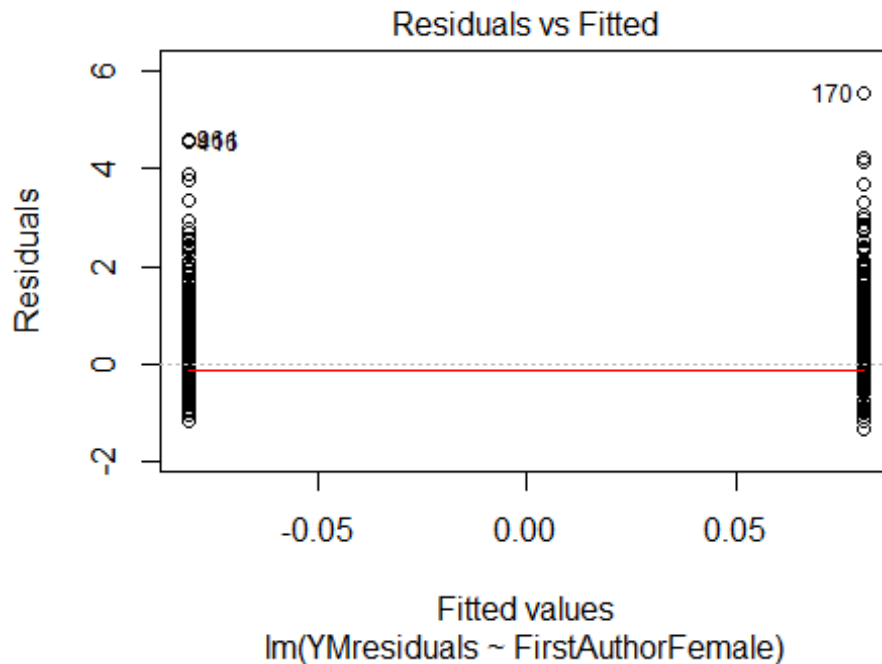
```



```
## 140 152
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 22 10 21 22 34 54 65 56 61 65 87 100 86 84
## 2011 2012
## 138 150
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 47, df = 16, p-value = 6e-05
```



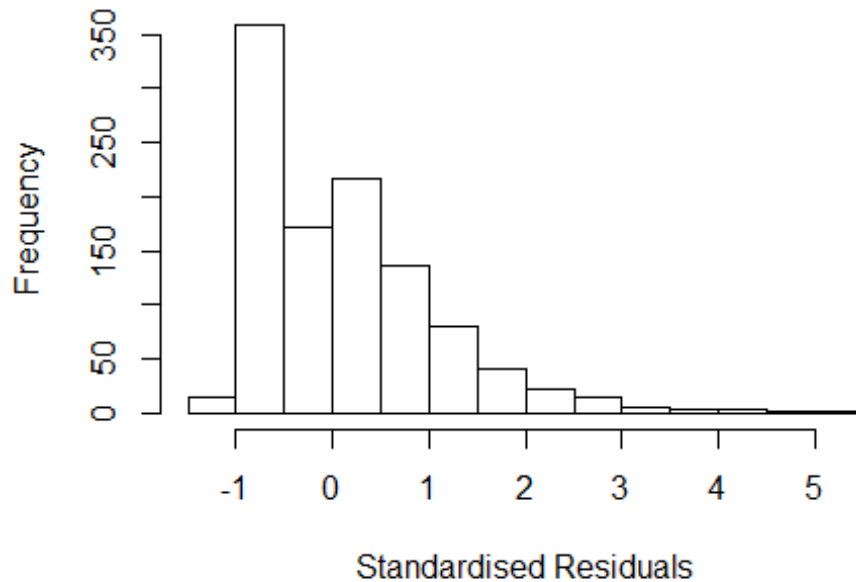
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.44, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 1.14107303380985"
## [1] "Male first author team size 2018 geometric mean: 1.16659659569439"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4100, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.16016175908666"
## [1] "Male last author team size 2018 geometric mean: 1.14010383540102"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, 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	1.854e+01	1	4.306e+00
## LastAuthorFemale	2.548e+13	1	5.048e+06
## UniqueAuthors	2.470e+14	4	6.296e+01
## Year	5.029e+13	16	2.680e+00

Residuals from first and last author and team size



```
## [1] "List of 28 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 167 33645282313 5.016 2002    1208      1    4.103
## 170 60949868527 6.843 2002    1208      2    5.499
## 293 70450081408 4.248 2003    1208      2    3.414
## 320 60950677030 5.223 2004    1208      1    4.274
## 327 60949299067 3.318 2004    1208      2    2.549
## 416 33750579723 5.330 2005    1208      1    4.474
## 515 67650070984 3.653 2006    1208      1    2.913
## 516 70450091776 4.126 2006    1208      1    3.206
## 538 60950108136 3.610 2007    1208      1    2.654
## 585 84870077833 3.610 2007    1208      1    2.834
## 618 68249150884 3.293 2007    1208      2    2.517
## 647 70449768665 3.802 2008    1208      1    2.842
## 684 67650076623 3.598 2008    1208      1    2.818
## 703 40449118934 3.462 2008    1202      4    2.682
## 784 70350630493 3.326 2009    1202      4    2.717
## 840 84976249888 3.306 2009    1208      1    2.517
## 919 77950240515 4.049 2010    1208      1    3.072
## 936 77950171071 3.832 2010    1208      1    3.035
## 961 84863442814 5.390 2011    1208      1    4.766
## 965 84855778105 4.563 2011    1208      2    3.939
## 1065 79960478515 3.818 2011    1208      2    3.014
## 1122 84868612124 4.382 2012    1208      1    3.585
## 1178 84866098127 4.858 2012    1208      2    3.882
## 1205 84862151897 3.937 2012    1208      1    2.709
## 1207 84864358305 5.374 2012    1208      1    4.398
```

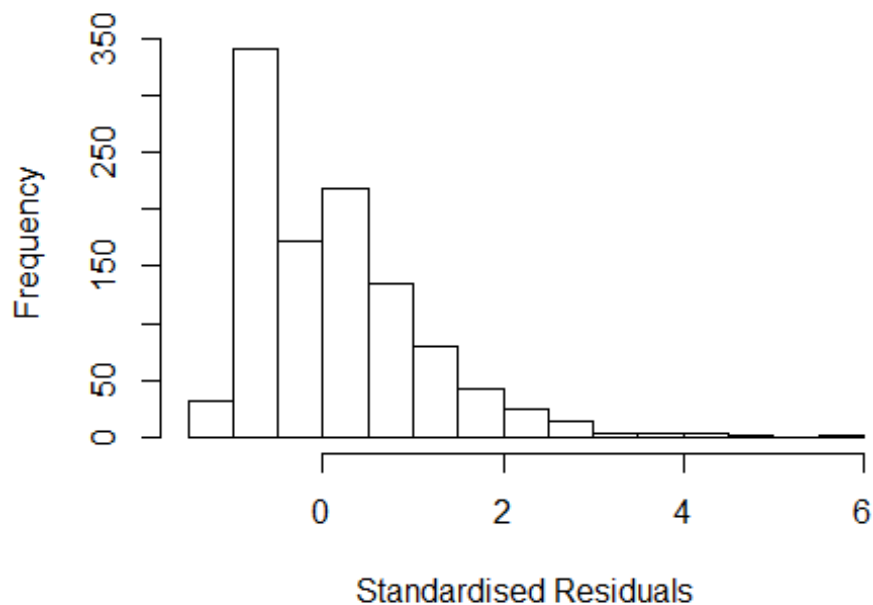
```

## 1217 84861928832 3.456 2012      1202      4      2.513
## 1231 84859348753 3.391 2012      1208      1      2.594
## 1232 84868281253 3.937 2012      1208      1      2.961
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2448 -0.7401 -0.0242  0.6524  5.4989
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5224    0.1761    2.97   0.0031 **
## FirstAuthorFemale1 0.0803    0.1721    0.47   0.6408
## LastAuthorFemale1 0.0993    0.1713    0.58   0.5621
## UniqueAuthors2    0.2517    0.1360    1.85   0.0645 .
## UniqueAuthors3    0.1466    0.2720    0.54   0.5900
## UniqueAuthors4    0.7001    0.0865    8.09 1.6e-15 ***
## UniqueAuthors5    1.1674    0.1835    6.36 3.0e-10 ***
## Year1997          0.0682    0.2161    0.32   0.7526
## Year1998          0.2391    0.2912    0.82   0.4119
## Year1999          0.2427    0.2250    1.08   0.2810
## Year2000          0.0107    0.2300    0.05   0.9628
## Year2001          0.0331    0.2170    0.15   0.8788
## Year2002          0.3903    0.2253    1.73   0.0835 .
## Year2003          0.1322    0.2071    0.64   0.5236
## Year2004          0.2466    0.2112    1.17   0.2432
## Year2005          0.0818    0.2034    0.40   0.6876
## Year2006          0.2177    0.2104    1.03   0.3011
## Year2007          0.2537    0.2086    1.22   0.2241
## Year2008          0.2575    0.1949    1.32   0.1866
## Year2009          0.0869    0.1944    0.45   0.6550
## Year2010          0.2747    0.1938    1.42   0.1567
## Year2011          0.1019    0.1908    0.53   0.5934
## Year2012          0.2744    0.1948    1.41   0.1592
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.858
## Multiple R-squared:  0.0292, Adjusted R-squared:  0.00876
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 6 observations c(125,128,248,326,788,997)
## are outliers with |weight| = 0 ( < 9.3e-05);

```

```
## 79 weights are ~= 1. The remaining 985 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0015 0.8900 0.9330 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      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 8.186 1 2.861
## LastAuthorFemale 8.193 1 2.862
## Year 1.050 16 1.002
```

Residuals from first and last author



```

## [1] "List of 27 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 167  33645282313 5.016 2002    1208      1    4.092
## 170  60949868527 6.843 2002    1208      2    5.735
## 293  70450081408 4.248 2003    1208      2    3.409
## 320  60950677030 5.223 2004    1208      1    4.262
## 327  60949299067 3.318 2004    1208      2    2.541
## 416  33750579723 5.330 2005    1208      1    4.711
## 515  67650070984 3.653 2006    1208      1    2.897
## 516  70450091776 4.126 2006    1208      1    3.186
## 538  60950108136 3.610 2007    1208      1    2.649
## 585  84870077833 3.610 2007    1208      1    2.833
## 618  68249150884 3.293 2007    1208      2    2.516
## 647  70449768665 3.802 2008    1208      1    2.830
## 684  67650076623 3.598 2008    1208      1    2.810
## 703  40449118934 3.462 2008    1202      4    2.674
## 784  70350630493 3.326 2009    1202      4    2.687
## 919  77950240515 4.049 2010    1208      1    3.065
## 936  77950171071 3.832 2010    1208      1    3.032
## 961  84863442814 5.390 2011    1208      1    4.735
## 965  84855778105 4.563 2011    1208      2    3.908
## 1065 79960478515 3.818 2011    1208      2    2.979
## 1122 84868612124 4.382 2012    1208      1    3.558
## 1178 84866098127 4.858 2012    1208      2    3.850
## 1205 84862151897 3.937 2012    1208      1    2.929
## 1207 84864358305 5.374 2012    1208      1    4.366
## 1217 84861928832 3.456 2012    1202      4    2.632
## 1231 84859348753 3.391 2012    1208      1    2.567
## 1232 84868281253 3.937 2012    1208      1    2.929
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1079 -0.7561 -0.0289  0.6587  5.7351
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.53961    0.17341   3.11  0.0019 **
## FirstAuthorFemale1 0.17979    0.15618   1.15  0.2499
## LastAuthorFemale1 0.00416    0.15654   0.03  0.9788
## Year1997        0.06640    0.21591   0.31  0.7585
## Year1998        0.22057    0.28978   0.76  0.4467
## Year1999        0.22343    0.22296   1.00  0.3165
## Year2000       -0.00225    0.22660  -0.01  0.9921
## Year2001        0.05061    0.21549   0.23  0.8144
## Year2002        0.38436    0.22306   1.72  0.0852 .

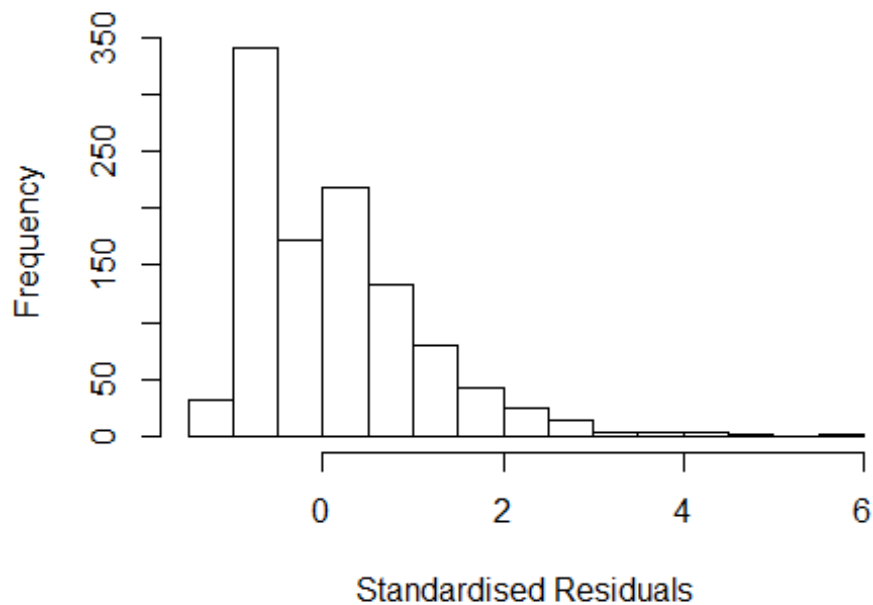
```

```

## Year2003      0.11590      0.20539      0.56      0.5727
## Year2004      0.23779      0.20936      1.14      0.2563
## Year2005      0.07979      0.20270      0.39      0.6939
## Year2006      0.21649      0.20715      1.05      0.2962
## Year2007      0.23713      0.20606      1.15      0.2501
## Year2008      0.24795      0.19289      1.29      0.1989
## Year2009      0.09968      0.19253      0.52      0.6048
## Year2010      0.26025      0.19210      1.35      0.1758
## Year2011      0.11498      0.18855      0.61      0.5421
## Year2012      0.28407      0.19344      1.47      0.1423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.859
## Multiple R-squared:  0.0221, Adjusted R-squared:  0.00539
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 6 observations c(125,128,248,326,788,997)
## are outliers with |weight| = 0 ( < 9.3e-05);
## 78 weights are ~= 1. The remaining 986 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0032 0.8890 0.9310 0.8990 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      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 1.014 1      1.007
## Year      1.014 16      1.000

```

Residuals from first author



```
## [1] "List of 27 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 167 33645282313 5.016 2002    1208      1    4.092
## 170 60949868527 6.843 2002    1208      2    5.735
## 293 70450081408 4.248 2003    1208      2    3.409
## 320 60950677030 5.223 2004    1208      1    4.262
## 327 60949299067 3.318 2004    1208      2    2.541
## 416 33750579723 5.330 2005    1208      1    4.711
## 515 67650070984 3.653 2006    1208      1    2.897
## 516 70450091776 4.126 2006    1208      1    3.186
## 538 60950108136 3.610 2007    1208      1    2.649
## 585 84870077833 3.610 2007    1208      1    2.833
## 618 68249150884 3.293 2007    1208      2    2.516
## 647 70449768665 3.802 2008    1208      1    2.830
## 684 67650076623 3.598 2008    1208      1    2.810
## 703 40449118934 3.462 2008    1202      4    2.674
## 784 70350630493 3.326 2009    1202      4    2.687
## 919 77950240515 4.049 2010    1208      1    3.065
## 936 77950171071 3.832 2010    1208      1    3.032
## 961 84863442814 5.390 2011    1208      1    4.735
## 965 84855778105 4.563 2011    1208      2    3.908
## 1065 79960478515 3.818 2011    1208      2    2.979
## 1122 84868612124 4.382 2012    1208      1    3.558
## 1178 84866098127 4.858 2012    1208      2    3.850
## 1205 84862151897 3.937 2012    1208      1    2.929
## 1207 84864358305 5.374 2012    1208      1    4.366
## 1217 84861928832 3.456 2012    1202      4    2.632
```



```

## 1231 84859348753 3.391 2012      1208      1      2.567
## 1232 84868281253 3.937 2012      1208      1      2.929
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.1075 -0.7559 -0.0289  0.6586  5.7355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.53963    0.17338   3.11  0.00191 **
## FirstAuthorFemale1 0.18374    0.05496   3.34  0.00086 ***
## Year1997        0.06664    0.21581   0.31  0.75756
## Year1998        0.22058    0.28981   0.76  0.44677
## Year1999        0.22346    0.22297   1.00  0.31648
## Year2000       -0.00236    0.22663  -0.01  0.99168
## Year2001        0.05067    0.21536   0.24  0.81404
## Year2002        0.38418    0.22304   1.72  0.08527 .
## Year2003        0.11562    0.20537   0.56  0.57355
## Year2004        0.23772    0.20937   1.14  0.25647
## Year2005        0.07958    0.20269   0.39  0.69470
## Year2006        0.21632    0.20715   1.04  0.29660
## Year2007        0.23685    0.20608   1.15  0.25069
## Year2008        0.24767    0.19283   1.28  0.19929
## Year2009        0.09935    0.19243   0.52  0.60578
## Year2010        0.26016    0.19211   1.35  0.17596
## Year2011        0.11469    0.18841   0.61  0.54285
## Year2012        0.28390    0.19343   1.47  0.14247
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.858
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.00635
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 6 observations c(125,128,248,326,788,997)
## are outliers with |weight| = 0 ( < 9.3e-05);
## 78 weights are ~= 1. The remaining 986 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.003  0.889   0.931   0.899   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 1.015 1          1.007
## Year            1.015 16          1.000

## [1] "List of 27 outliers with residuals above 2.5"
##          ScopusId  NLCS Year OneField Fields residuals
## 167 33645282313 5.016 2002 1208 1 4.092
## 170 60949868527 6.843 2002 1208 2 5.735
## 293 70450081408 4.248 2003 1208 2 3.409
## 320 60950677030 5.223 2004 1208 1 4.262
## 327 60949299067 3.318 2004 1208 2 2.541
## 416 33750579723 5.330 2005 1208 1 4.711
## 515 67650070984 3.653 2006 1208 1 2.897
## 516 70450091776 4.126 2006 1208 1 3.186
## 538 60950108136 3.610 2007 1208 1 2.649
## 585 84870077833 3.610 2007 1208 1 2.833
## 618 68249150884 3.293 2007 1208 2 2.516
## 647 70449768665 3.802 2008 1208 1 2.830
## 684 67650076623 3.598 2008 1208 1 2.810
## 703 40449118934 3.462 2008 1202 4 2.674
## 784 70350630493 3.326 2009 1202 4 2.687
## 919 77950240515 4.049 2010 1208 1 3.065
## 936 77950171071 3.832 2010 1208 1 3.032
## 961 84863442814 5.390 2011 1208 1 4.735
## 965 84855778105 4.563 2011 1208 2 3.908
## 1065 79960478515 3.818 2011 1208 2 2.979
## 1122 84868612124 4.382 2012 1208 1 3.558
## 1178 84866098127 4.858 2012 1208 2 3.850
## 1205 84862151897 3.937 2012 1208 1 2.929
## 1207 84864358305 5.374 2012 1208 1 4.366
## 1217 84861928832 3.456 2012 1202 4 2.632
## 1231 84859348753 3.391 2012 1208 1 2.567
## 1232 84868281253 3.937 2012 1208 1 2.929
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))

```

```

## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1065 -0.7589 -0.0349  0.6555  5.7365
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.54555    0.17368   3.14  0.0017 **
## LastAuthorFemale1 0.17220    0.05518   3.12  0.0019 **
## Year1997        0.05707    0.21520   0.27  0.7909
## Year1998        0.21836    0.28953   0.75  0.4509
## Year1999        0.22318    0.22307   1.00  0.3173
## Year2000        0.00571    0.22559   0.03  0.9798
## Year2001        0.04761    0.21685   0.22  0.8263
## Year2002        0.38877    0.22314   1.74  0.0818 .
## Year2003        0.11762    0.20610   0.57  0.5683
## Year2004        0.23792    0.20863   1.14  0.2544
## Year2005        0.08184    0.20342   0.40  0.6875
## Year2006        0.21331    0.20777   1.03  0.3048
## Year2007        0.23761    0.20633   1.15  0.2498
## Year2008        0.25159    0.19324   1.30  0.1932
## Year2009        0.10596    0.19293   0.55  0.5830
## Year2010        0.26057    0.19230   1.36  0.1757
## Year2011        0.12216    0.18892   0.65  0.5180
## Year2012        0.28109    0.19365   1.45  0.1469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.859
## Multiple R-squared:  0.0209, Adjusted R-squared:  0.00506
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 6 observations c(125,128,248,326,788,997)
## are outliers with |weight| = 0 ( < 9.3e-05);
## 81 weights are ~= 1. The remaining 983 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0039 0.8900 0.9300 0.8990 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          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: 1070"
## [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]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    1    1    3    1    1    4    3   16   23   16   23   23   24   23
## 2012
##    26
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    1    1    3    1    1    4    3   12   19   11   19   20   23   21
## 2012
##    23
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    1    1    3    1    1    4    3   11   19   11   18   20   22   21
## 2012
##    22
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.17346046000463"
## [1] "Male first author team size 2018 geometric mean: 1.53746267757577"

## 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 = 43, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.28088668976427"
## [1] "Male last author team size 2018 geometric mean: 1.36426160182137"

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

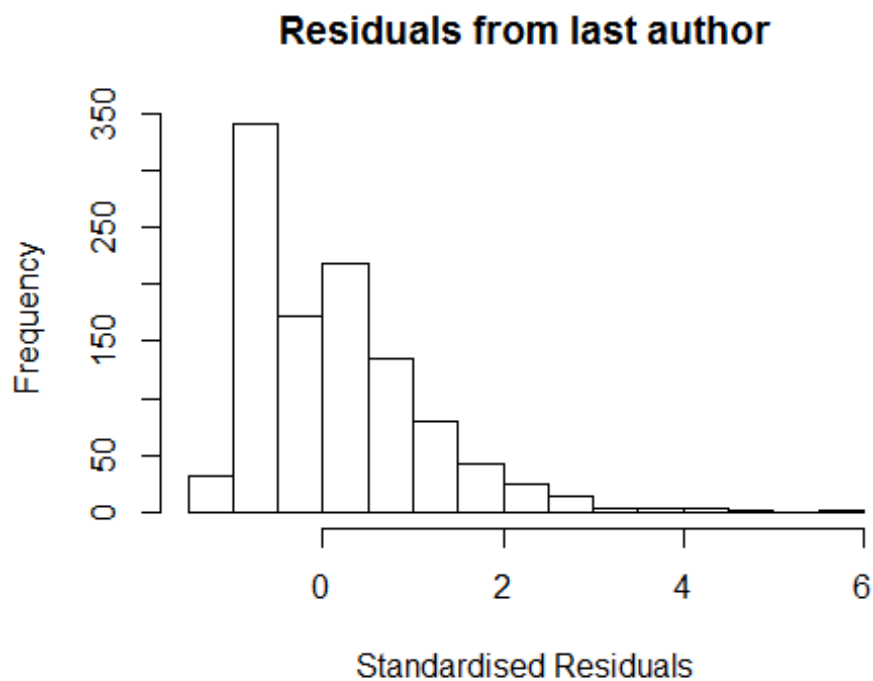
##
## Wilcoxon rank sum test with continuity correction
##

```

```
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 51, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

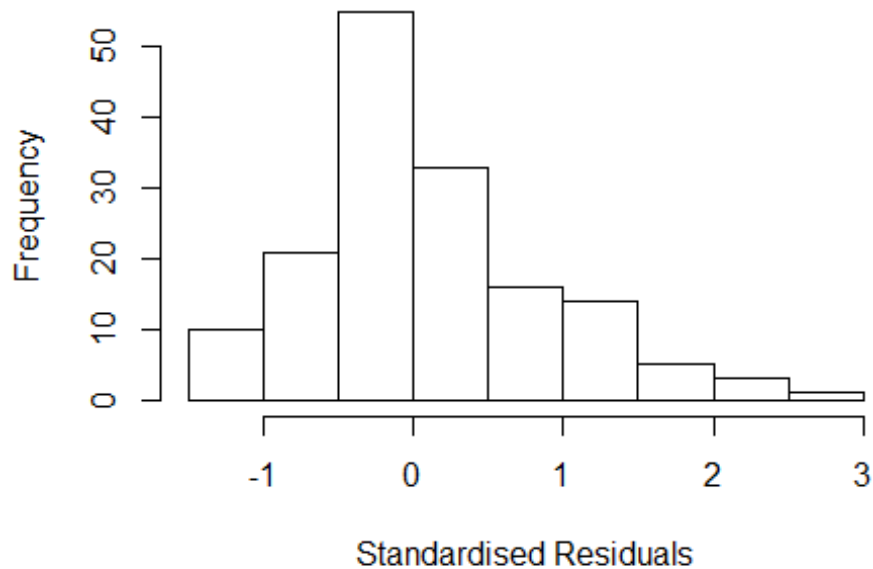
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## 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
## LastAuthorFemale	NaN	1	NaN
## UniqueAuthors	NaN	4	NaN
## Year	NaN	14	NaN

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
## 209 85011526508 2.938 2012    1202    5    2.623
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3727 -0.4919 -0.0336  0.4837  2.6229
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2000     0.0000      Inf < 2e-16 ***
## FirstAuthorFemale1 -0.2615     0.2237    -1.17  0.24447
## LastAuthorFemale1  0.0815     0.2052     0.40  0.69181
## UniqueAuthors2     0.0100     0.1996     0.05  0.96011
## UniqueAuthors3     0.9781     0.2322     4.21  4.5e-05 ***
## UniqueAuthors4     1.2110     0.3034     3.99  0.00011 ***
## UniqueAuthors5    -0.0355     0.2159    -0.16  0.86978
## Year1999          -0.0060     0.0000    -Inf < 2e-16 ***
## Year2000          -0.6602     0.1733    -3.81  0.00021 ***
## Year2001          -0.6621     0.1692    -3.91  0.00014 ***
```

```

## Year2002          -1.2000      0.0000      -Inf < 2e-16 ***
## Year2003          -0.6701      0.3196      -2.10 0.03785 *
## Year2004           0.1515      0.1525       0.99 0.32200
## Year2005          -0.0192      0.2358      -0.08 0.93537
## Year2006          -0.0199      0.2016      -0.10 0.92151
## Year2007           0.1594      0.2464       0.65 0.51882
## Year2008           0.2815      0.2159       1.30 0.19453
## Year2009          -0.6796      0.2383      -2.85 0.00502 **
## Year2010          -0.1025      0.2483      -0.41 0.68036
## Year2011           0.1627      0.3144       0.52 0.60557
## Year2012          -0.7050      0.3385      -2.08 0.03914 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.686
## Multiple R-squared:  0.293, Adjusted R-squared:  0.19
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 141 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.112  0.840   0.952   0.879   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      6.33e-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 2.751e+13 1      5.245e+06
## LastAuthorFemale  2.419e+13 1      4.918e+06
## Year              3.156e+13 14      3.035e+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
## -1.5538 -0.5638 -0.0823  0.6323  2.4894
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.20e+00   1.56e-07  7.68e+06 < 2e-16 ***
## FirstAuthorFemale1 -4.41e-01   2.11e-01 -2.09e+00  0.03881 *
## LastAuthorFemale1  2.47e-01   1.98e-01  1.24e+00  0.21540
## Year1999         -6.00e-03   1.39e-07 -4.33e+04 < 2e-16 ***
## Year2000         -5.93e-01   1.37e-01 -4.34e+00  2.7e-05 ***
## Year2001         -6.48e-01   1.76e-01 -3.68e+00  0.00033 ***
## Year2002         -1.20e+00   1.45e-07 -8.27e+06 < 2e-16 ***
## Year2003         -6.11e-01   5.54e-01 -1.10e+00  0.27176
## Year2004          4.80e-01   1.62e-01  2.95e+00  0.00370 **
## Year2005         -1.75e-02   2.19e-01 -8.00e-02  0.93634
## Year2006          6.13e-02   2.07e-01  3.00e-01  0.76748
## Year2007          1.69e-01   2.24e-01  7.50e-01  0.45253
## Year2008          3.44e-01   1.84e-01  1.87e+00  0.06332 .
## Year2009         -6.29e-01   2.72e-01 -2.32e+00  0.02202 *
## Year2010         -1.82e-02   2.48e-01 -7.00e-02  0.94154
## Year2011          3.54e-01   2.86e-01  1.24e+00  0.21739
## Year2012         -5.57e-01   3.85e-01 -1.45e+00  0.14990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.706
## Multiple R-squared:  0.213, Adjusted R-squared:  0.123
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 146 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.189  0.830  0.938  0.879  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      6.33e-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

```

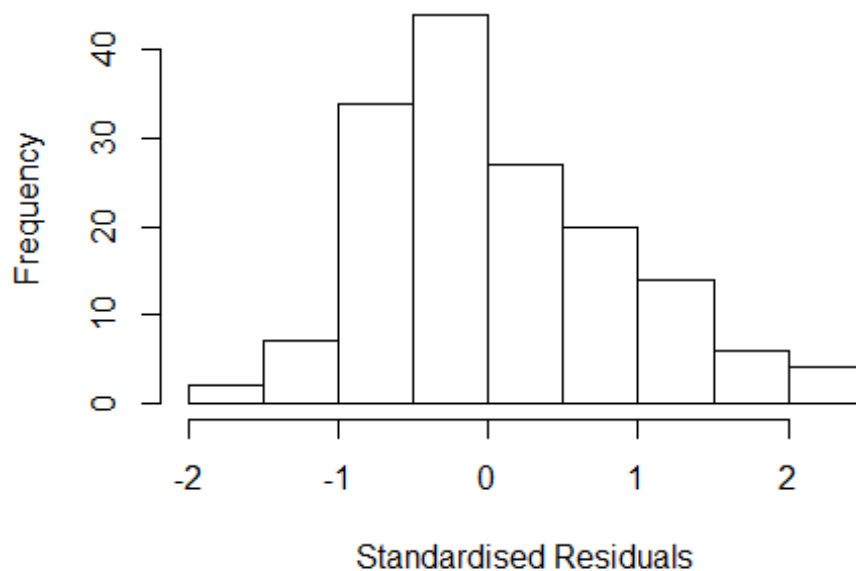


```
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## 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 14 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
```

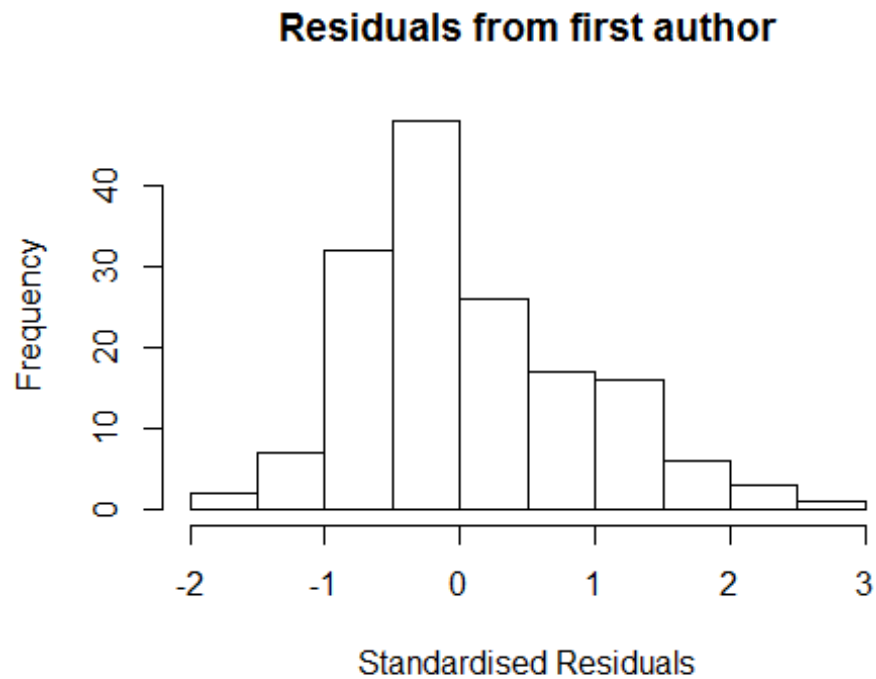
```

## -1.5776 -0.5323 -0.0685 0.6169 2.5201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2000     0.0000      Inf < 2e-16 ***
## FirstAuthorFemale1 -0.2538     0.1669     -1.52 0.13052
## Year1999          -0.0060     0.0000     -Inf < 2e-16 ***
## Year2000          -0.6350     0.1765     -3.60 0.00044 ***
## Year2001          -0.5882     0.1669     -3.52 0.00057 ***
## Year2002          -1.2000     0.0000     -Inf < 2e-16 ***
## Year2003          -0.6289     0.5604     -1.12 0.26367
## Year2004           0.4995     0.1516      3.29 0.00125 **
## Year2005           0.0168     0.2314      0.07 0.94237
## Year2006           0.0823     0.2070      0.40 0.69156
## Year2007           0.1812     0.2224      0.81 0.41668
## Year2008           0.4007     0.1772      2.26 0.02525 *
## Year2009          -0.5987     0.2696     -2.22 0.02798 *
## Year2010          -0.0067     0.2359     -0.03 0.97737
## Year2011           0.3776     0.2973      1.27 0.20603
## Year2012          -0.5283     0.3965     -1.33 0.18490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.697
## Multiple R-squared: 0.209, Adjusted R-squared: 0.125
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 143 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.164 0.834 0.933 0.873 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      6.33e-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"

```

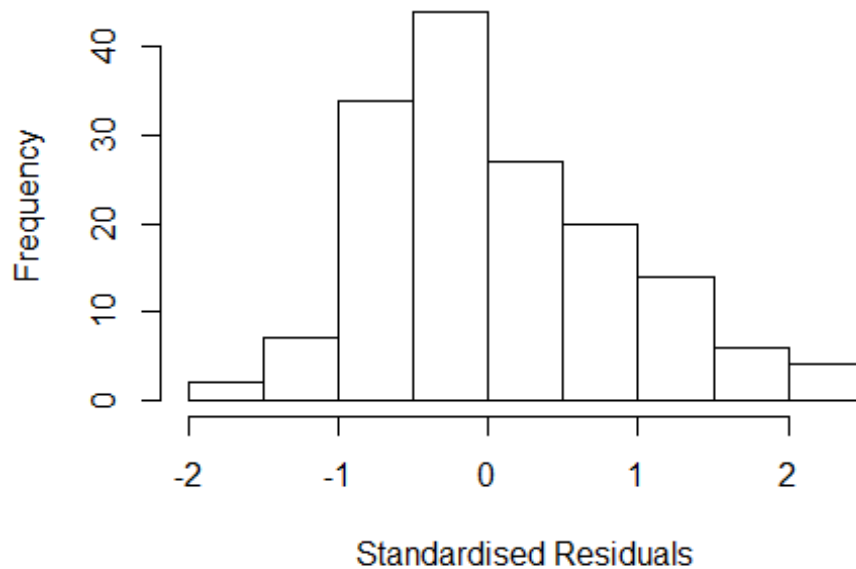
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## 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 14          NaN
```

Residuals from last author



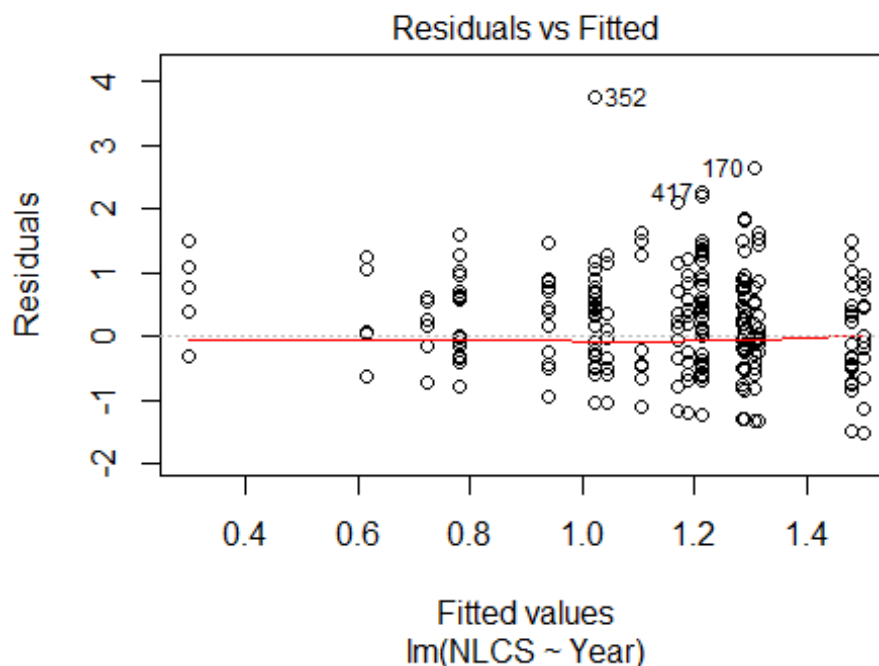
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4562 -0.4998 -0.0719 0.5761 2.3849
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2000 0.0000 Inf < 2e-16 ***
## LastAuthorFemale1 -0.0787 0.1598 -0.49 0.6232
## Year1999 -0.0060 0.0000 -Inf < 2e-16 ***
## Year2000 -0.7750 0.1861 -4.16 5.4e-05 ***
## Year2001 -0.7633 0.1598 -4.77 4.4e-06 ***
## Year2002 -1.2000 0.0000 -Inf < 2e-16 ***
## Year2003 -0.6002 0.5328 -1.13 0.2619
## Year2004 0.4411 0.1827 2.41 0.0170 *
## Year2005 -0.0510 0.2496 -0.20 0.8385
## Year2006 -0.0229 0.1855 -0.12 0.9018
## Year2007 0.1420 0.2263 0.63 0.5314
## Year2008 0.3066 0.1914 1.60 0.1114
```

```

## Year2009          -0.7002      0.2656    -2.64    0.0093 **
## Year2010          -0.1424      0.2181    -0.65    0.5149
## Year2011           0.2561      0.3078     0.83    0.4067
## Year2012          -0.5682      0.3900    -1.46    0.1474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.188, Adjusted R-squared:  0.103
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 143 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.226  0.805   0.939   0.873   0.970   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 158"
## [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
##   13   16   21    7   14   11   15   14   20   24   26   38   20   23   33
## 2011 2012
##   47   66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12   14   19    7   14   11   14   10   17   18   17   33   17   20   27
## 2011 2012
##   42   61
##

```

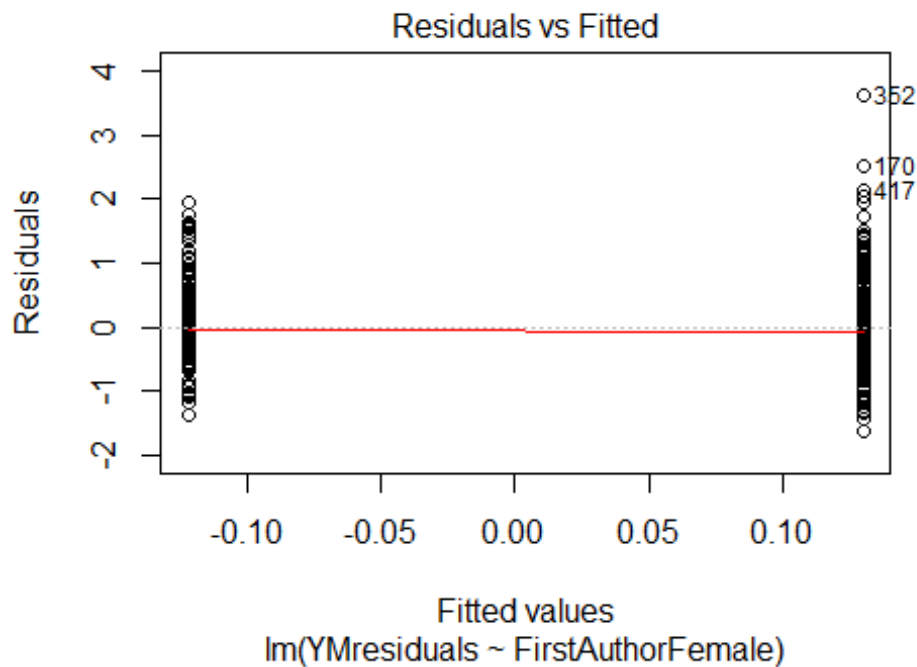
```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    12    14    19     7    14    11    14    10    17    18    17    33    16    20    26
## 2011 2012
##    41    59
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 13, df = 16, p-value = 0.7
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9, df = 1, p-value = 0.003
## [1] "Female first author team size 2018 geometric mean: 1.5140248023509"
## [1] "Male first author team size 2018 geometric mean: 1.7043533238261"
## 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 = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.37574755138729"
## [1] "Male last author team size 2018 geometric mean: 1.83381975637347"

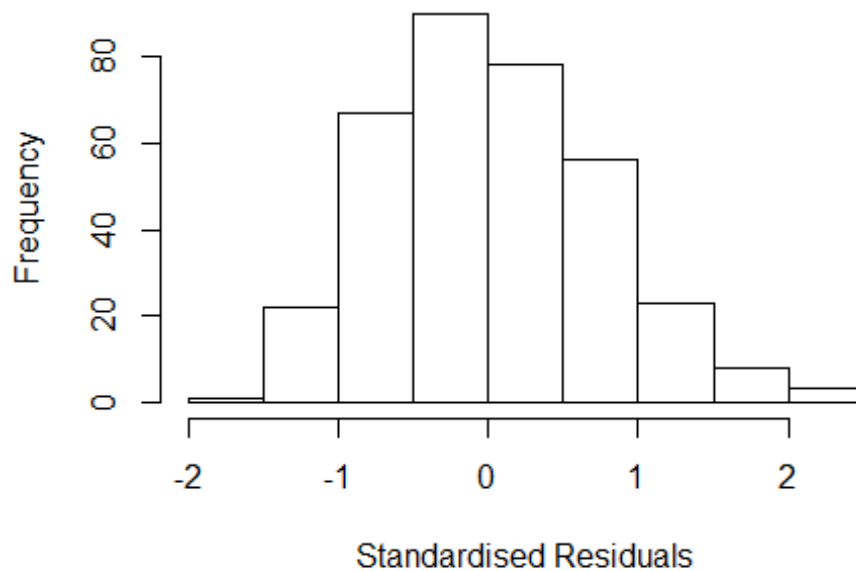
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	4.018	1	2.004
LastAuthorFemale	3.673	1	1.917
UniqueAuthors	2.029	4	1.092
Year	2.346	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
## -1.6594 -0.5092 -0.0126 0.5292 2.2489
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9181 0.3225 2.85 0.0047 **
## FirstAuthorFemale1 0.1523 0.1529 1.00 0.3199
## LastAuthorFemale1 -0.0593 0.1502 -0.39 0.6935
## UniqueAuthors2 0.5606 0.1103 5.08 6.3e-07 ***
## UniqueAuthors3 0.7987 0.1797 4.45 1.2e-05 ***
## UniqueAuthors4 1.1517 0.3632 3.17 0.0017 **
## UniqueAuthors5 1.4855 0.4061 3.66 0.0003 ***
## Year1997 0.3542 0.3975 0.89 0.3736
## Year1998 -0.7043 0.3436 -2.05 0.0412 *
## Year1999 -0.4056 0.3715 -1.09 0.2757
```

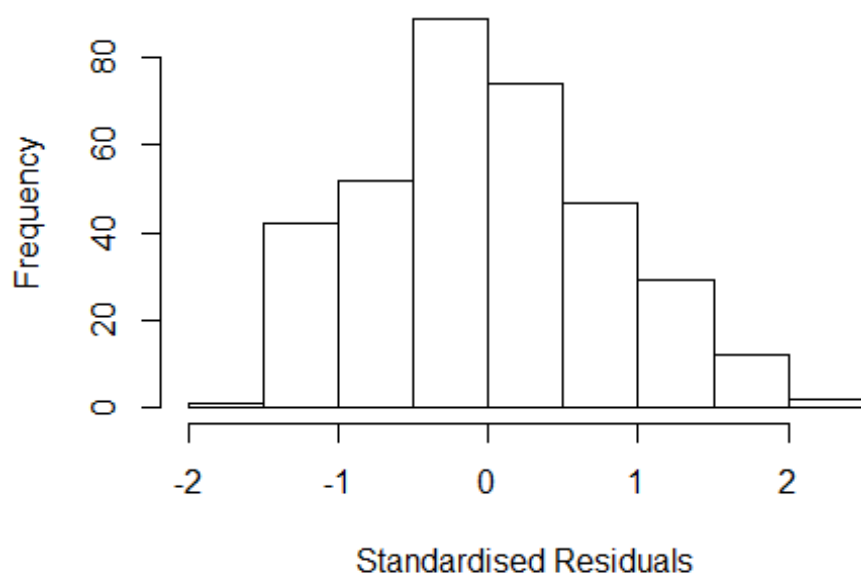


```

## Year2000      0.0177      0.3666      0.05      0.9616
## Year2001     -0.3908      0.4109     -0.95      0.3422
## Year2002     -0.1849      0.3710     -0.50      0.6186
## Year2003     -0.1400      0.4515     -0.31      0.7566
## Year2004      0.2175      0.3898      0.56      0.5773
## Year2005     -0.2522      0.3663     -0.69      0.4916
## Year2006      0.0449      0.3621      0.12      0.9015
## Year2007      0.1807      0.3820      0.47      0.6365
## Year2008      0.1004      0.3691      0.27      0.7857
## Year2009      0.1725      0.3701      0.47      0.6415
## Year2010     -0.5650      0.3555     -1.59      0.1129
## Year2011     -0.2689      0.3480     -0.77      0.4402
## Year2012     -0.0823      0.3506     -0.23      0.8147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.734
## Multiple R-squared:  0.276, Adjusted R-squared:  0.227
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.327  0.870  0.949  0.909  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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.800 1      1.949
## LastAuthorFemale  3.838 1      1.959
## Year              1.459 16      1.012

```

Residuals from first and last author



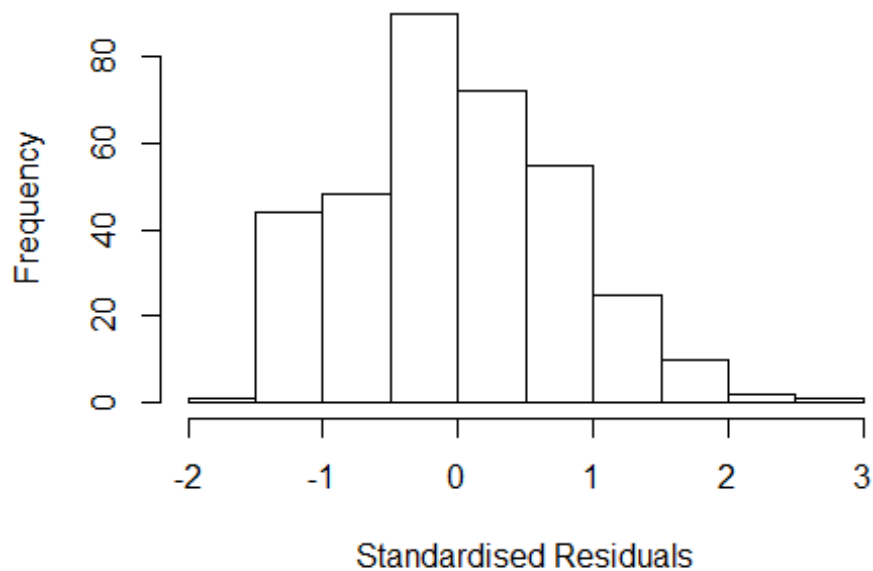
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5358 -0.5240 -0.0636 0.5383 2.4068
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0764 0.2956 3.64 0.00031 ***
## FirstAuthorFemale1 0.3972 0.1656 2.40 0.01704 *
## LastAuthorFemale1 -0.1965 0.1703 -1.15 0.24935
## Year1997 0.4141 0.3646 1.14 0.25689
## Year1998 -0.8795 0.3221 -2.73 0.00667 **
## Year1999 -0.4610 0.3512 -1.31 0.19020
## Year2000 0.0772 0.3812 0.20 0.83959
## Year2001 -0.5524 0.3860 -1.43 0.15339
## Year2002 -0.2261 0.3753 -0.60 0.54737
## Year2003 -0.2130 0.5003 -0.43 0.67060
## Year2004 0.1342 0.3871 0.35 0.72905
## Year2005 -0.2911 0.3534 -0.82 0.41072
```

```

## Year2006          0.0546      0.3569      0.15  0.87859
## Year2007          0.0790      0.3510      0.23  0.82197
## Year2008          0.0569      0.3694      0.15  0.87777
## Year2009          0.2587      0.3494      0.74  0.45967
## Year2010         -0.4379      0.3351     -1.31  0.19222
## Year2011         -0.2698      0.3273     -0.82  0.41035
## Year2012         -0.0405      0.3308     -0.12  0.90251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.81
## Multiple R-squared:  0.147, Adjusted R-squared:  0.0998
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 319 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.357  0.864  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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.201 1      1.096
## Year              1.201 16      1.006

```

Residuals from first author



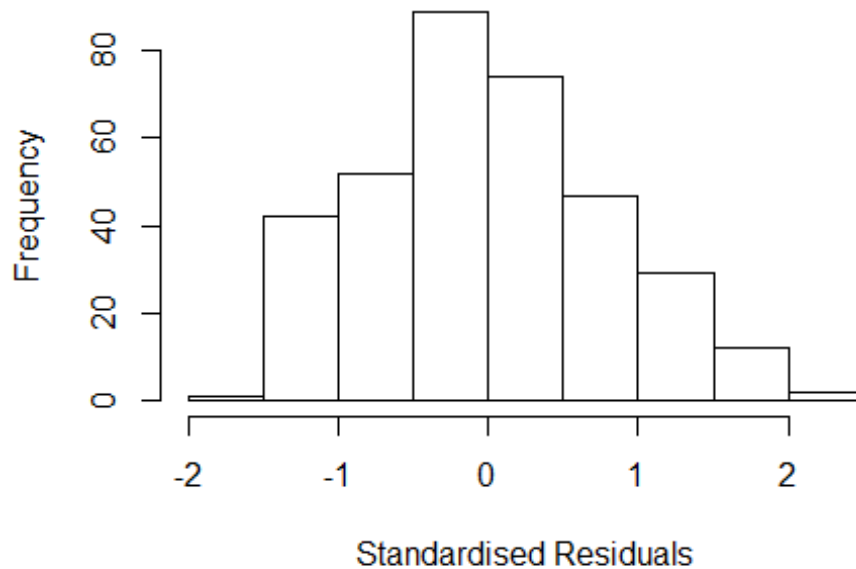
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5638 -0.5155 -0.0655 0.5541 2.6114
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07201 0.29505 3.63 0.00032 ***
## FirstAuthorFemale1 0.24714 0.09412 2.63 0.00905 **
## Year1997 0.40053 0.36073 1.11 0.26766
## Year1998 -0.89003 0.32060 -2.78 0.00581 **
## Year1999 -0.44834 0.35425 -1.27 0.20655
## Year2000 0.05875 0.37702 0.16 0.87627
## Year2001 -0.55663 0.38393 -1.45 0.14805
## Year2002 -0.22663 0.37850 -0.60 0.54975
## Year2003 -0.23578 0.50279 -0.47 0.63943
## Year2004 0.13548 0.38404 0.35 0.72449
## Year2005 -0.27510 0.35744 -0.77 0.44206
## Year2006 0.00445 0.34756 0.01 0.98979
```

```

## Year2007          0.05511      0.34941      0.16  0.87477
## Year2008          0.02816      0.36523      0.08  0.93859
## Year2009          0.24468      0.34944      0.70  0.48430
## Year2010         -0.45313      0.32988     -1.37  0.17049
## Year2011         -0.26892      0.32706     -0.82  0.41154
## Year2012         -0.05246      0.32875     -0.16  0.87332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.807
## Multiple R-squared:  0.141, Adjusted R-squared:  0.0967
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.273  0.860  0.948  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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.283 1      1.133
## Year              1.283 16      1.008

```

Residuals from last author



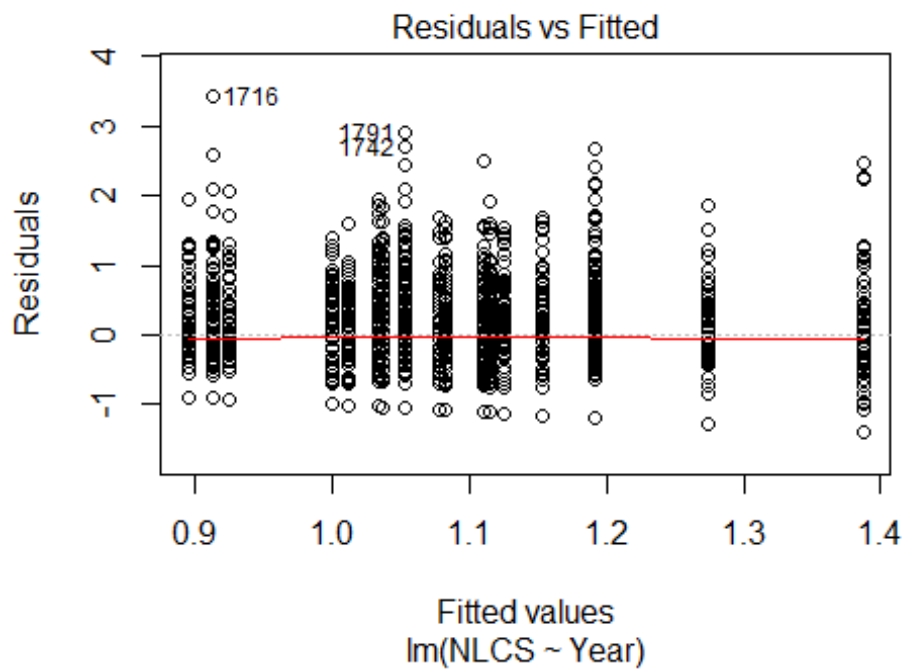
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5102 -0.5105 -0.0683 0.5902 2.8183
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0834 0.2978 3.64 0.00032 ***
## LastAuthorFemale1 0.1101 0.0982 1.12 0.26296
## Year1997 0.4196 0.3629 1.16 0.24843
## Year1998 -0.8582 0.3243 -2.65 0.00853 **
## Year1999 -0.3847 0.3648 -1.05 0.29243
## Year2000 0.0745 0.3753 0.20 0.84285
## Year2001 -0.5443 0.3906 -1.39 0.16439
## Year2002 -0.1336 0.3827 -0.35 0.72728
## Year2003 -0.1719 0.4952 -0.35 0.72880
## Year2004 0.1744 0.3898 0.45 0.65480
## Year2005 -0.1978 0.3595 -0.55 0.58249
## Year2006 0.0333 0.3500 0.10 0.92415
```

```

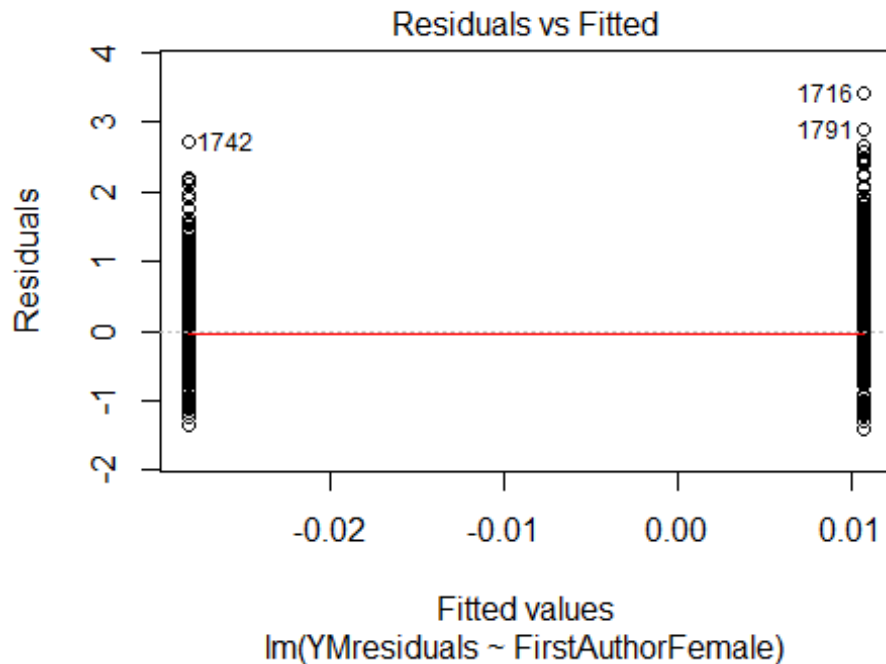
## Year2007          0.1232      0.3549      0.35  0.72883
## Year2008          0.0949      0.3727      0.25  0.79917
## Year2009          0.3167      0.3570      0.89  0.37571
## Year2010         -0.4087      0.3323     -1.23  0.21951
## Year2011         -0.2010      0.3286     -0.61  0.54124
## Year2012          0.0145      0.3322      0.04  0.96515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.807
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0806
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 322 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.198  0.867   0.953   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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 348"
## [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
##   67   77   96   98  101  105   86   98  121  105  121  126  152  169  133
## 2011 2012
##  182  210
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   55   61   79   80   79   80   66   80  100   92  112  106  127  145  120
## 2011 2012

```

```
## 165 189
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 60 78 79 77 78 66 79 99 89 111 106 125 139 118
## 2011 2012
## 163 186
## [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
```

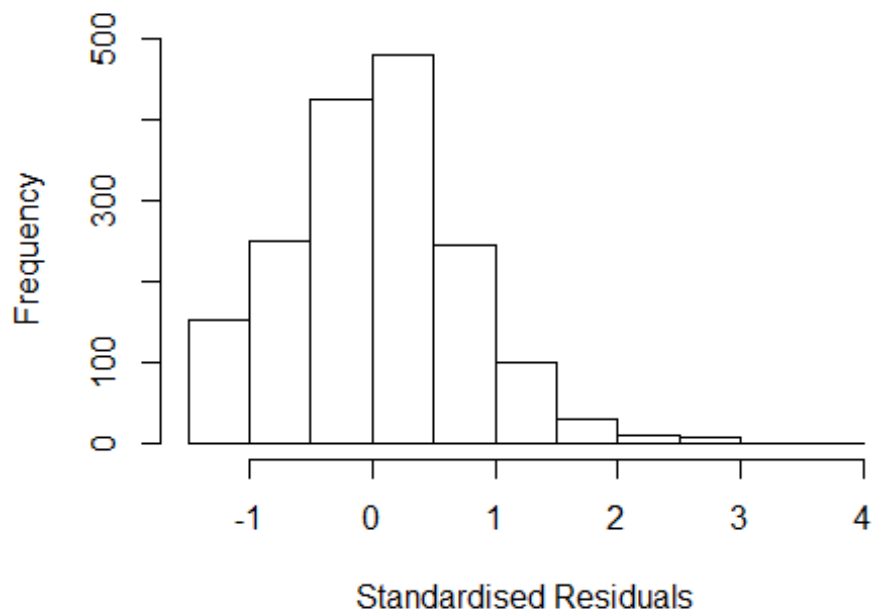


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.93, df = 1, p-value = 0.3
```

```
## [1] "Female first author team size 2018 geometric mean: 1.38814679442775"
## [1] "Male first author team size 2018 geometric mean: 1.18189866004016"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5600, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.4738497128933"
## [1] "Male last author team size 2018 geometric mean: 1.13641949496734"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6200, 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 4.021 1          2.005
## LastAuthorFemale  4.084 1          2.021
## UniqueAuthors    1.328 4          1.036
## Year              1.225 16         1.006
```

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
## 1230 42049108384 3.623 2007    1211      1    2.553
## 1688 77950007656 3.500 2010    1211      1    2.672
## 1716 72249083366 4.339 2010    1211      1    3.511
## 1742 80053033678 3.746 2011    1211      1    2.834
## 1745 81855182200 3.481 2011    1211      1    2.511
## 1791 80053164465 3.974 2011    1211      1    3.004
## 1935 84871220174 3.867 2012    1211      1    2.778
## 2090 84859110805 3.600 2012    1211      1    2.511
## 2146 84857934111 3.867 2012    1211      1    2.778
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4677 -0.4734  0.0206  0.4630  3.5108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0572     0.0993   10.65 < 2e-16 ***
## FirstAuthorFemale1 -0.0104     0.0778   -0.13  0.8934
```

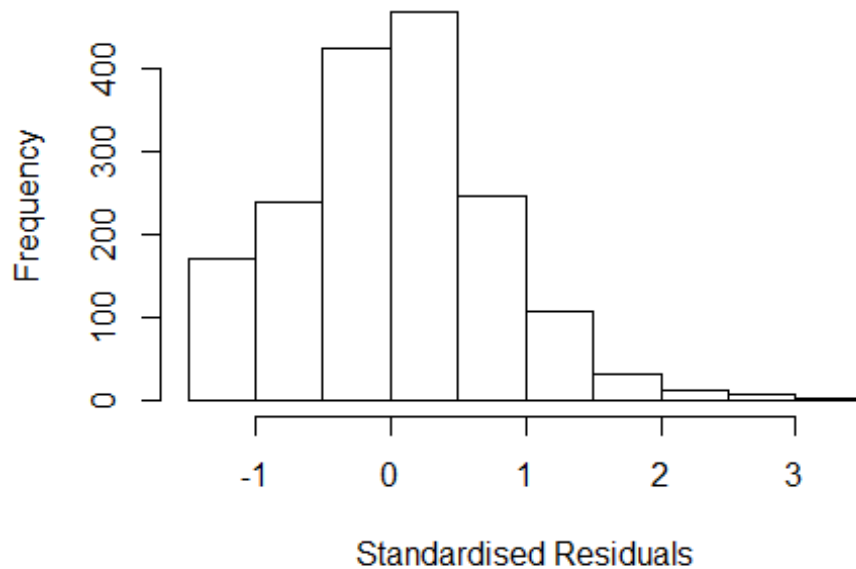
```

## LastAuthorFemale1 -0.0473 0.0770 -0.61 0.5389
## UniqueAuthors2 0.1455 0.0666 2.19 0.0290 *
## UniqueAuthors3 0.5081 0.1094 4.65 3.7e-06 ***
## UniqueAuthors4 0.5015 0.1727 2.90 0.0037 **
## UniqueAuthors5 0.4670 0.2822 1.65 0.0981 .
## Year1997 -0.1522 0.1342 -1.13 0.2569
## Year1998 -0.2017 0.1253 -1.61 0.1077
## Year1999 -0.0556 0.1288 -0.43 0.6659
## Year2000 -0.0556 0.1253 -0.44 0.6573
## Year2001 0.0455 0.1224 0.37 0.7104
## Year2002 0.2062 0.1285 1.60 0.1088
## Year2003 -0.0438 0.1312 -0.33 0.7383
## Year2004 0.0586 0.1222 0.48 0.6316
## Year2005 0.2875 0.1248 2.30 0.0214 *
## Year2006 -0.0115 0.1208 -0.10 0.9239
## Year2007 0.0128 0.1183 0.11 0.9141
## Year2008 0.0357 0.1151 0.31 0.7565
## Year2009 -0.0774 0.1150 -0.67 0.5009
## Year2010 -0.2290 0.1173 -1.95 0.0510 .
## Year2011 -0.0873 0.1201 -0.73 0.4674
## Year2012 0.0318 0.1215 0.26 0.7935
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.705
## Multiple R-squared: 0.0445, Adjusted R-squared: 0.032
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1340 is an outlier with |weight| = 0 ( < 5.9e-05);
## 177 weights are ~= 1. The remaining 1528 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0296 0.8440 0.9490 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 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 4.675  1      2.162
## LastAuthorFemale  4.752  1      2.180
## Year              1.086 16      1.003
```

Residuals from first and last author



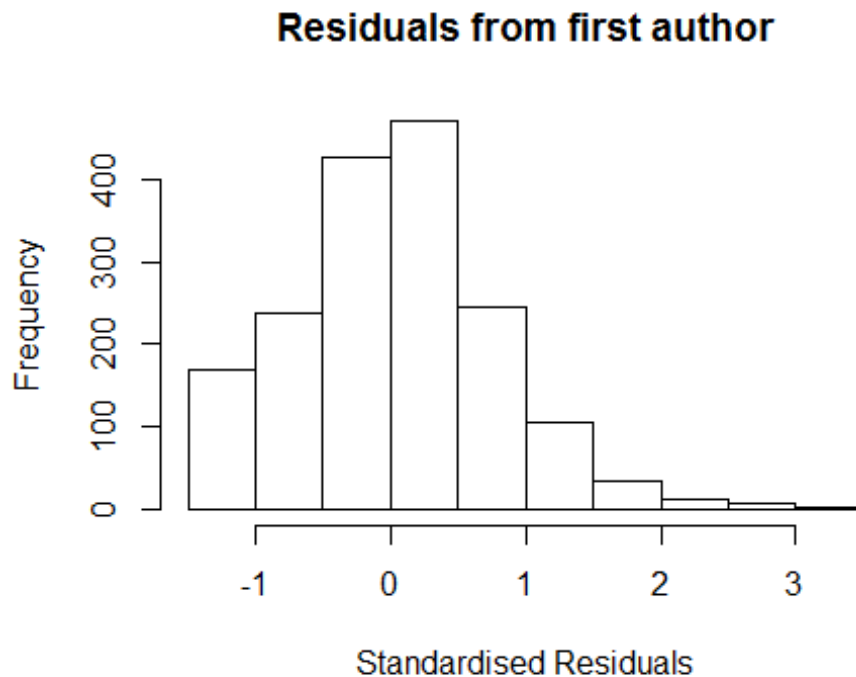
```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1230 42049108384 3.623 2007    1211      1      2.528
## 1688 77950007656 3.500 2010    1211      1      2.657
## 1716 72249083366 4.339 2010    1211      1      3.496
## 1742 80053033678 3.746 2011    1211      1      2.782
## 1791 80053164465 3.974 2011    1211      1      2.974
## 1935 84871220174 3.867 2012    1211      1      2.729
## 2146 84857934111 3.867 2012    1211      1      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
## -1.3571 -0.4766  0.0189  0.4684  3.4961
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept)          1.0586      0.0994    10.65   <2e-16 ***
## FirstAuthorFemale1    0.0219      0.0850     0.26    0.796
## LastAuthorFemale1    -0.0581      0.0848    -0.69    0.493
## Year1997              -0.1540      0.1346    -1.14    0.253
## Year1998              -0.1969      0.1250    -1.58    0.115
## Year1999              -0.0382      0.1289    -0.30    0.767
## Year2000              -0.0200      0.1261    -0.16    0.874
## Year2001               0.0483      0.1234     0.39    0.695
## Year2002               0.2193      0.1295     1.69    0.090 .
## Year2003              -0.0390      0.1308    -0.30    0.765
## Year2004               0.0839      0.1224     0.69    0.493
## Year2005               0.2985      0.1245     2.40    0.017 *
## Year2006               0.0218      0.1216     0.18    0.857
## Year2007               0.0366      0.1183     0.31    0.757
## Year2008               0.0624      0.1158     0.54    0.590
## Year2009              -0.0557      0.1155    -0.48    0.630
## Year2010              -0.2157      0.1181    -1.83    0.068 .
## Year2011              -0.0588      0.1203    -0.49    0.625
## Year2012               0.0790      0.1211     0.65    0.514
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.713
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.0178
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1340 is an outlier with |weight| = 0 ( < 5.9e-05);
## 162 weights are ~= 1. The remaining 1543 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0427 0.8450 0.9510 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      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 1.032 1 1.016
## Year 1.032 16 1.001
```



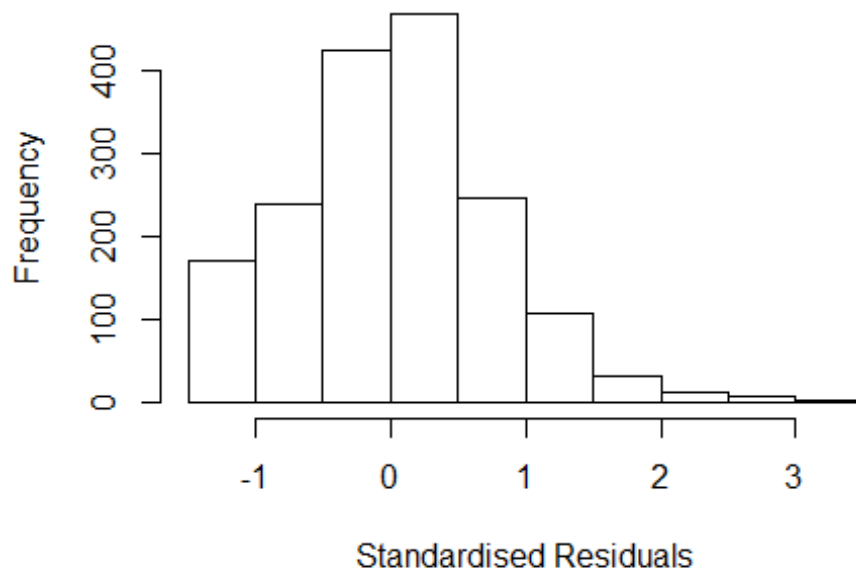
```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1230 42049108384 3.623 2007    1211      1    2.528
## 1688 77950007656 3.500 2010    1211      1    2.657
## 1716 72249083366 4.339 2010    1211      1    3.496
## 1742 80053033678 3.746 2011    1211      1    2.782
## 1791 80053164465 3.974 2011    1211      1    2.974
## 1935 84871220174 3.867 2012    1211      1    2.729
## 2146 84857934111 3.867 2012    1211      1    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
## -1.3535 -0.4784  0.0181  0.4630  3.4961
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0565     0.0998   10.59  <2e-16 ***
## FirstAuthorFemale1 -0.0290     0.0400   -0.73    0.468
## Year1997       -0.1535     0.1350   -1.14    0.256
```

```

## Year1998          -0.1957      0.1253    -1.56      0.119
## Year1999          -0.0388      0.1292    -0.30      0.764
## Year2000          -0.0224      0.1261    -0.18      0.859
## Year2001           0.0491      0.1236      0.40      0.691
## Year2002           0.2180      0.1297      1.68      0.093 .
## Year2003          -0.0381      0.1312    -0.29      0.771
## Year2004           0.0837      0.1227      0.68      0.495
## Year2005           0.2970      0.1248      2.38      0.017 *
## Year2006           0.0206      0.1219      0.17      0.866
## Year2007           0.0388      0.1187      0.33      0.744
## Year2008           0.0631      0.1161      0.54      0.587
## Year2009          -0.0560      0.1159    -0.48      0.629
## Year2010          -0.2136      0.1183    -1.81      0.071 .
## Year2011          -0.0598      0.1205    -0.50      0.620
## Year2012           0.0787      0.1215      0.65      0.517
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0278, Adjusted R-squared:  0.018
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1340 is an outlier with |weight| = 0 ( < 5.9e-05);
## 162 weights are ~= 1. The remaining 1543 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0415 0.8430 0.9510 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      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 1.049 1          1.024
## Year            1.049 16          1.001

```

Residuals from last author



```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1230 42049108384 3.623 2007    1211      1    2.528
## 1688 77950007656 3.500 2010    1211      1    2.657
## 1716 72249083366 4.339 2010    1211      1    3.496
## 1742 80053033678 3.746 2011    1211      1    2.782
## 1791 80053164465 3.974 2011    1211      1    2.974
## 1935 84871220174 3.867 2012    1211      1    2.729
## 2146 84857934111 3.867 2012    1211      1    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
## -1.3576 -0.4744  0.0184  0.4690  3.4946
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0588     0.0994   10.65  <2e-16 ***
## LastAuthorFemale1 -0.0389     0.0398   -0.98   0.329
## Year1997         -0.1537     0.1346   -1.14   0.254
## Year1998         -0.1969     0.1251   -1.57   0.116
## Year1999         -0.0381     0.1289   -0.30   0.768
## Year2000         -0.0203     0.1260   -0.16   0.872
```

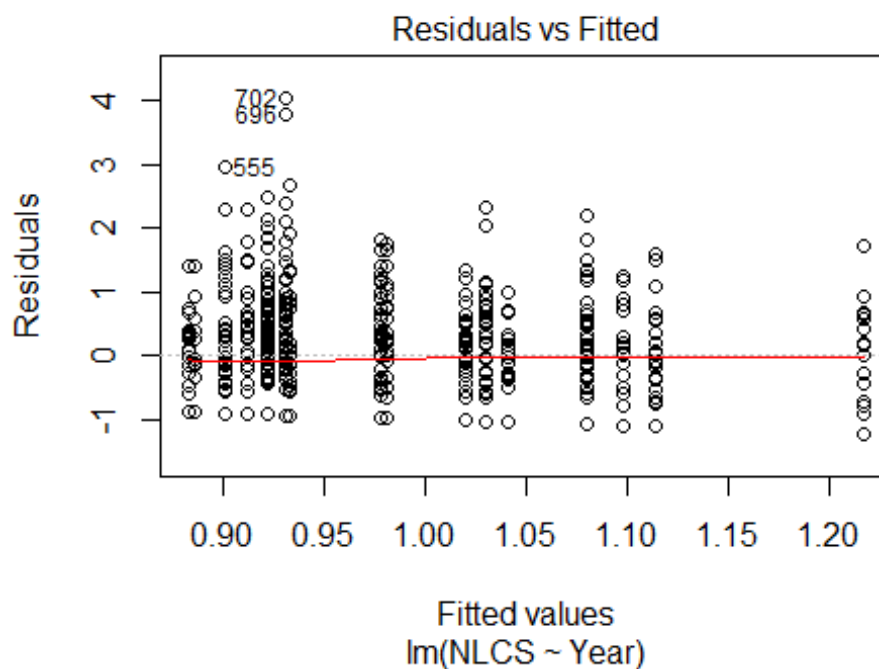


```

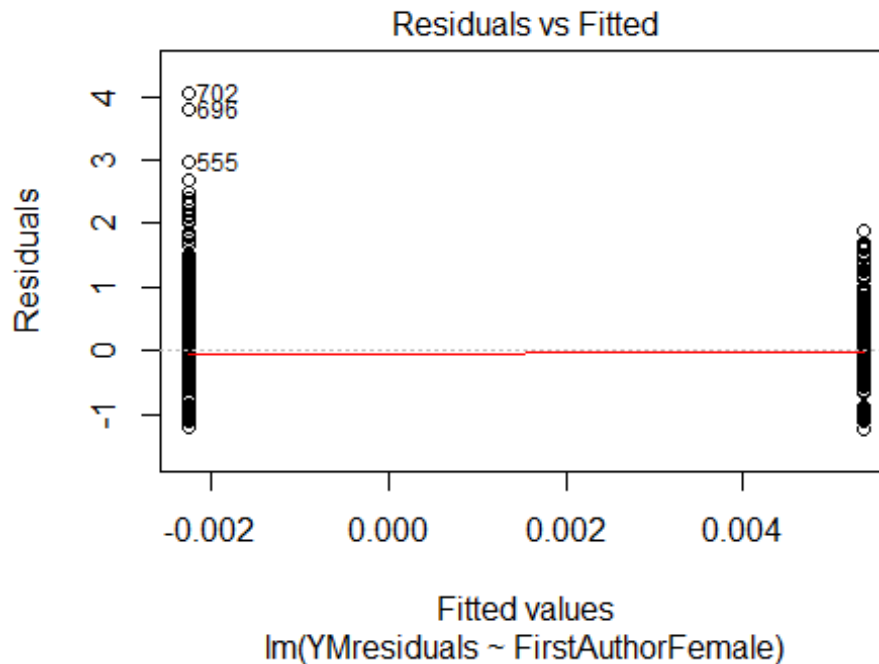
## Year2001      0.0491      0.1233      0.40      0.690
## Year2002      0.2196      0.1295      1.70      0.090 .
## Year2003     -0.0384      0.1309     -0.29      0.769
## Year2004      0.0844      0.1224      0.69      0.491
## Year2005      0.2988      0.1245      2.40      0.017 *
## Year2006      0.0222      0.1217      0.18      0.855
## Year2007      0.0381      0.1183      0.32      0.747
## Year2008      0.0630      0.1158      0.54      0.586
## Year2009     -0.0550      0.1157     -0.48      0.634
## Year2010     -0.2143      0.1180     -1.82      0.069 .
## Year2011     -0.0586      0.1204     -0.49      0.626
## Year2012      0.0791      0.1212      0.65      0.514
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.0183
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1340 is an outlier with |weight| = 0 ( < 5.9e-05);
## 160 weights are ~ = 1. The remaining 1545 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0423 0.8450 0.9510 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           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: 1706"
## [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

```

```
## 18 27 25 32 26 22 47 35 45 36 41 71 52 64 60
## 2011 2012
## 109 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 25 22 30 22 18 42 33 39 35 38 58 47 55 50
## 2011 2012
## 98 91
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 25 22 30 22 18 42 33 38 35 38 57 47 53 49
## 2011 2012
## 97 91
## [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
```

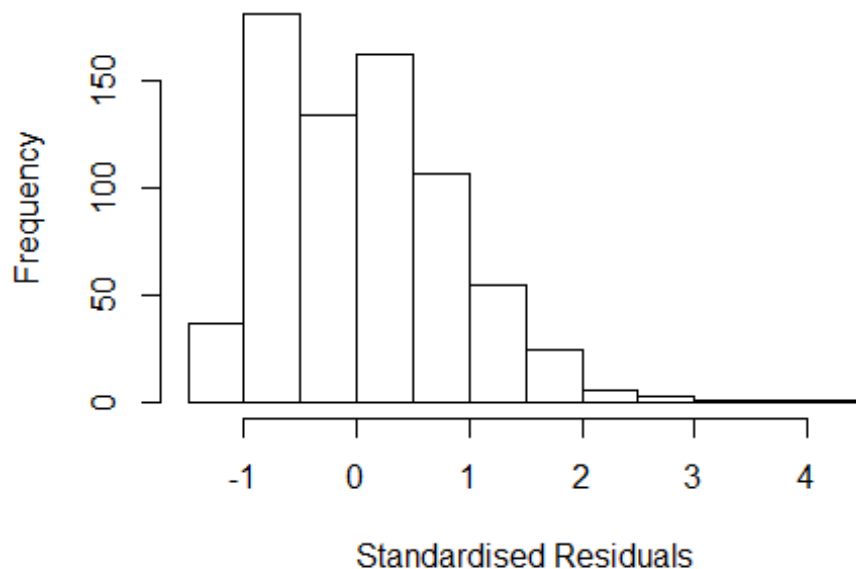


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.2, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 1.5585067515696"
## [1] "Male first author team size 2018 geometric mean: 1.1870129089269"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.59991277356716"
## [1] "Male last author team size 2018 geometric mean: 1.14987855762878"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.001
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 5.268 1          2.295
## LastAuthorFemale  5.092 1          2.257
## UniqueAuthors    1.773 4          1.074
## Year             1.701 16          1.017
```

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
## 89  37249083928 3.610 1999    1212     1    2.825
## 555 84890571282 3.880 2010    1212     1    3.167
## 693 79960482431 3.313 2011    1202     2    2.580
## 696 79952203580 4.710 2011    1212     1    3.600
## 702 84875391890 4.974 2011    1212     1    4.241
## 773 84862327953 3.424 2012    1202     2    2.628
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.33603 -0.73305  0.00843  0.58083  4.24095
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82142    0.19716   4.17 3.5e-05 ***
## FirstAuthorFemale1 -0.12015    0.14518  -0.83  0.408
## LastAuthorFemale1  0.18350    0.14731   1.25  0.213
## UniqueAuthors2     0.31633    0.15447   2.05  0.041 *
## UniqueAuthors3     0.37713    0.09174   4.11 4.4e-05 ***
```

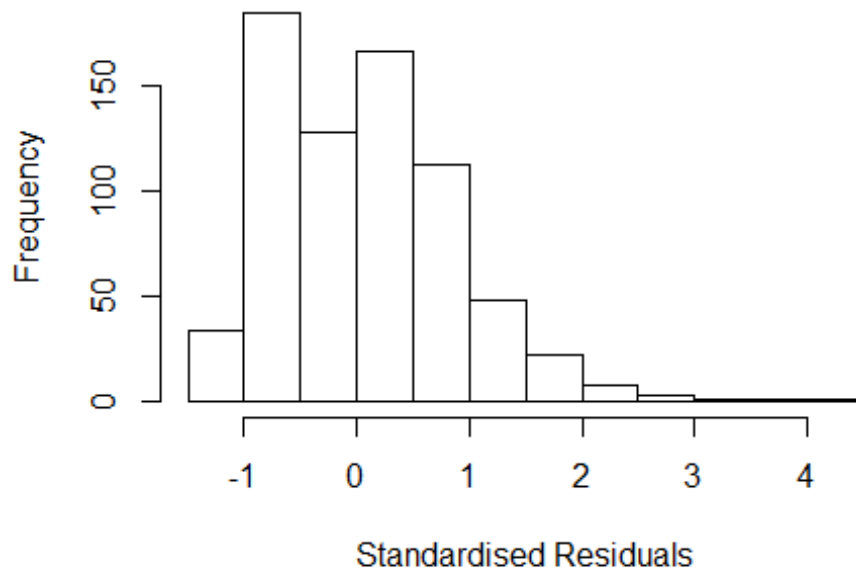
```

## UniqueAuthors4      0.73055      0.65725      1.11      0.267
## UniqueAuthors5      0.41150      0.39091      1.05      0.293
## Year1997             0.22381      0.23435      0.96      0.340
## Year1998             0.25181      0.24917      1.01      0.313
## Year1999            -0.03643      0.24930     -0.15      0.884
## Year2000            -0.00190      0.22976     -0.01      0.993
## Year2001             0.37071      0.27441      1.35      0.177
## Year2002            -0.02404      0.24333     -0.10      0.921
## Year2003             0.25593      0.23588      1.09      0.278
## Year2004             0.00133      0.23709      0.01      0.996
## Year2005             0.06425      0.24838      0.26      0.796
## Year2006             0.13492      0.22900      0.59      0.556
## Year2007             0.19680      0.22304      0.88      0.378
## Year2008             0.12339      0.22751      0.54      0.588
## Year2009             0.04787      0.22567      0.21      0.832
## Year2010            -0.10890      0.23387     -0.47      0.642
## Year2011            -0.08838      0.21375     -0.41      0.679
## Year2012            -0.02499      0.21761     -0.11      0.909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.827
## Multiple R-squared:  0.0368, Adjusted R-squared:  0.00602
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 599 is an outlier with |weight| = 0 ( < 0.00014);
## 56 weights are ~= 1. The remaining 654 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0188 0.8980 0.9340 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      1.41e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.072 1          2.018

```

```
## LastAuthorFemale 4.002 1 2.001
## Year 1.183 16 1.005
```

Residuals from first and last author



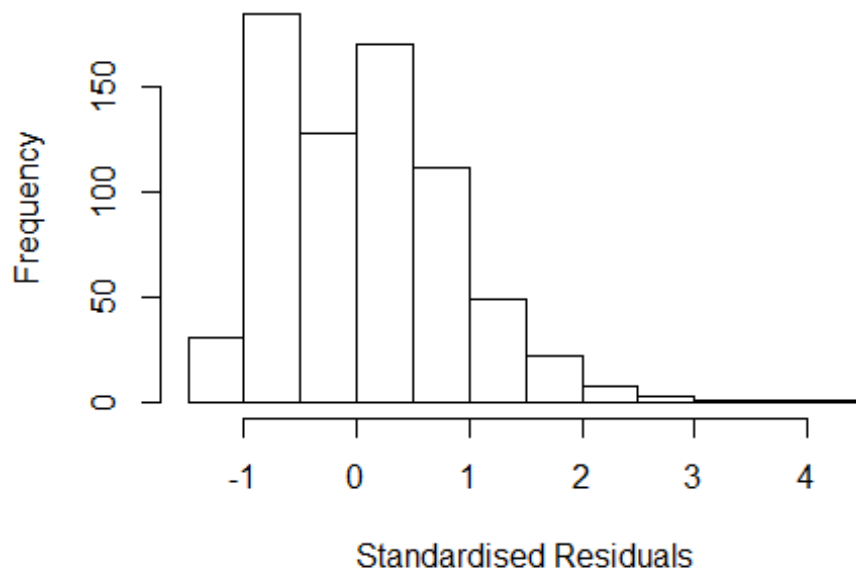
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 89  37249083928 3.610 1999    1212     1    2.811
## 555 84890571282 3.880 2010    1212     1    3.119
## 693 79960482431 3.313 2011    1202     2    2.545
## 696 79952203580 4.710 2011    1212     1    3.942
## 702 84875391890 4.974 2011    1212     1    4.206
## 773 84862327953 3.424 2012    1202     2    2.573
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.269 -0.768  0.035  0.568  4.206
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.86174    0.20424   4.22 2.8e-05 ***
## FirstAuthorFemale1 -0.04347    0.12866  -0.34  0.74
## LastAuthorFemale1  0.12358    0.13108   0.94  0.35
## Year1997         0.18225    0.24020   0.76  0.45
```

```

## Year1998      0.20552    0.25401    0.81    0.42
## Year1999     -0.06319    0.25372   -0.25    0.80
## Year2000     -0.02023    0.23480   -0.09    0.93
## Year2001      0.32740    0.27961    1.17    0.24
## Year2002     -0.03582    0.24718   -0.14    0.88
## Year2003      0.22012    0.24110    0.91    0.36
## Year2004      0.00182    0.24560    0.01    0.99
## Year2005      0.07192    0.25630    0.28    0.78
## Year2006      0.12360    0.23674    0.52    0.60
## Year2007      0.17255    0.22964    0.75    0.45
## Year2008      0.09423    0.23496    0.40    0.69
## Year2009      0.04120    0.23297    0.18    0.86
## Year2010     -0.10100    0.24087   -0.42    0.68
## Year2011     -0.09416    0.22124   -0.43    0.67
## Year2012     -0.01093    0.22428   -0.05    0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.857
## Multiple R-squared:  0.0204, Adjusted R-squared:  -0.00513
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 599 is an outlier with |weight| = 0 ( < 0.00014);
## 58 weights are ~= 1. The remaining 652 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0012 0.8950 0.9370 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      1.41e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.121 1          1.059
## Year              1.121 16          1.004

```

Residuals from first author



```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 89  37249083928 3.610 1999    1212     1    2.811
## 555 84890571282 3.880 2010    1212     1    3.119
## 693 79960482431 3.313 2011    1202     2    2.545
## 696 79952203580 4.710 2011    1212     1    3.942
## 702 84875391890 4.974 2011    1212     1    4.206
## 773 84862327953 3.424 2012    1202     2    2.573
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2557 -0.7687  0.0421  0.5752  4.2053
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.85815    0.20119   4.27 2.3e-05 ***
## FirstAuthorFemale1 0.06302    0.06752   0.93  0.35
## Year1997        0.18670    0.23756   0.79  0.43
## Year1998        0.21532    0.25221   0.85  0.39
## Year1999       -0.05734    0.25147  -0.23  0.82
## Year2000       -0.00816    0.23320  -0.03  0.97
## Year2001        0.33457    0.27703   1.21  0.23
```

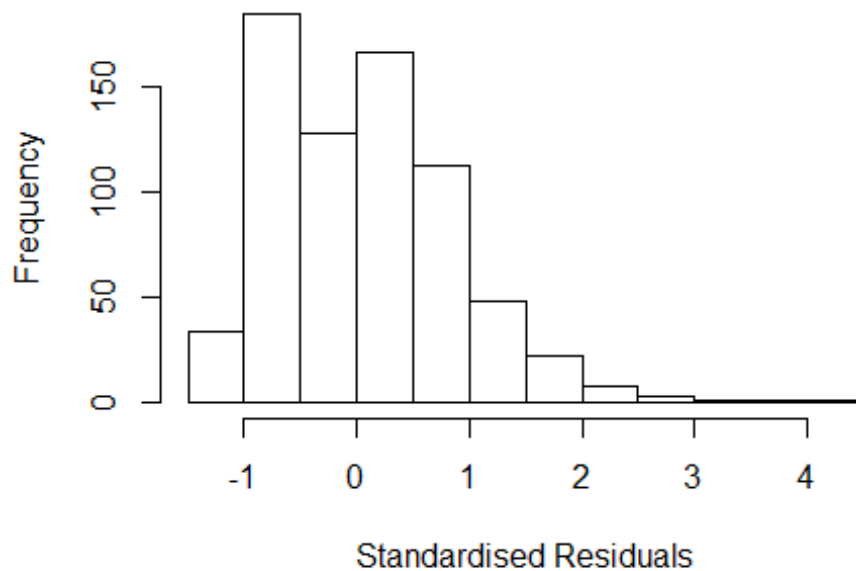


```

## Year2002      -0.03076    0.24509   -0.13    0.90
## Year2003      0.22388    0.23886    0.94    0.35
## Year2004      0.01763    0.24364    0.07    0.94
## Year2005      0.07879    0.25446    0.31    0.76
## Year2006      0.13320    0.23397    0.57    0.57
## Year2007      0.17610    0.22718    0.78    0.44
## Year2008      0.10038    0.23239    0.43    0.67
## Year2009      0.04598    0.23036    0.20    0.84
## Year2010     -0.09691    0.23889   -0.41    0.69
## Year2011     -0.08943    0.21868   -0.41    0.68
## Year2012     -0.00453    0.22169   -0.02    0.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.849
## Multiple R-squared:  0.0195, Adjusted R-squared:  -0.00456
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 599 is an outlier with |weight| = 0 ( < 0.00014);
## 55 weights are ~= 1. The remaining 655 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0004 0.8970 0.9360 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          1.41e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.104 1          1.051
## Year            1.104 16          1.003

```

Residuals from last author



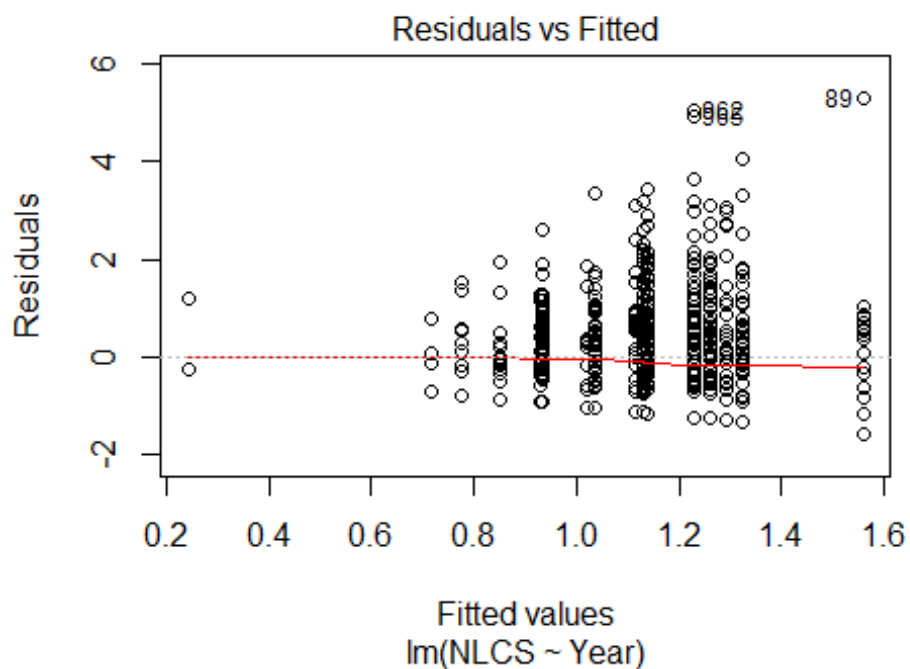
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 89  37249083928 3.610 1999    1212     1    2.811
## 555 84890571282 3.880 2010    1212     1    3.119
## 693 79960482431 3.313 2011    1202     2    2.545
## 696 79952203580 4.710 2011    1212     1    3.942
## 702 84875391890 4.974 2011    1212     1    4.206
## 773 84862327953 3.424 2012    1202     2    2.573
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.273 -0.764  0.036  0.570  4.210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.85795    0.20322   4.22 2.7e-05 ***
## LastAuthorFemale1 0.08560    0.06899   1.24  0.22
## Year1997        0.18596    0.23942   0.78  0.44
## Year1998        0.20726    0.25317   0.82  0.41
## Year1999       -0.06329    0.25293  -0.25  0.80
## Year2000       -0.01849    0.23424  -0.08  0.94
## Year2001        0.32978    0.27923   1.18  0.24
```

```

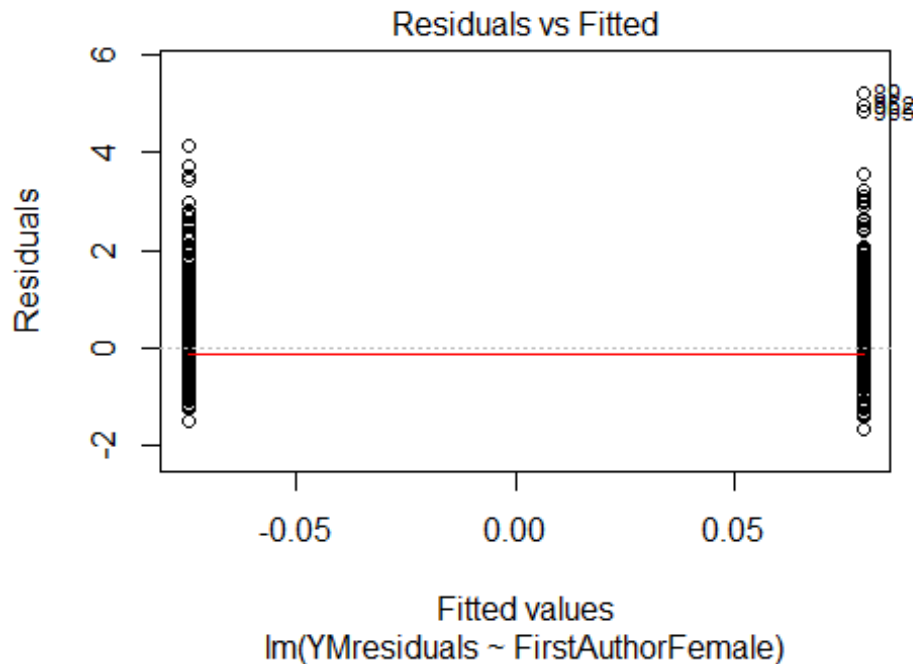
## Year2002      -0.03620      0.24642      -0.15      0.88
## Year2003      0.22106      0.24048      0.92      0.36
## Year2004      0.00433      0.24539      0.02      0.99
## Year2005      0.07343      0.25604      0.29      0.77
## Year2006      0.12523      0.23604      0.53      0.60
## Year2007      0.17204      0.22865      0.75      0.45
## Year2008      0.09370      0.23398      0.40      0.69
## Year2009      0.04239      0.23193      0.18      0.86
## Year2010     -0.10245      0.23996     -0.43      0.67
## Year2011     -0.09435      0.22028     -0.43      0.67
## Year2012     -0.01070      0.22342     -0.05      0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.847
## Multiple R-squared:  0.0205, Adjusted R-squared:  -0.00353
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(594,599)
## are outliers with |weight| <= 0.00014 ( < 0.00014);
## 61 weights are ~= 1. The remaining 648 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.145  0.893  0.935  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      1.41e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 711"
## [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

```

```
##      6      6      20      13      15      23      26      48      44      42      46      82      115      86      117
## 2011 2012
## 159 184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      4      6      16      13      12      22      23      41      42      37      40      74      100      77      102
## 2011 2012
## 135 157
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      4      6      16      13      11      22      23      41      41      37      40      74      99      77      98
## 2011 2012
## 130 155
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 64, df = 16, p-value = 9e-08
```

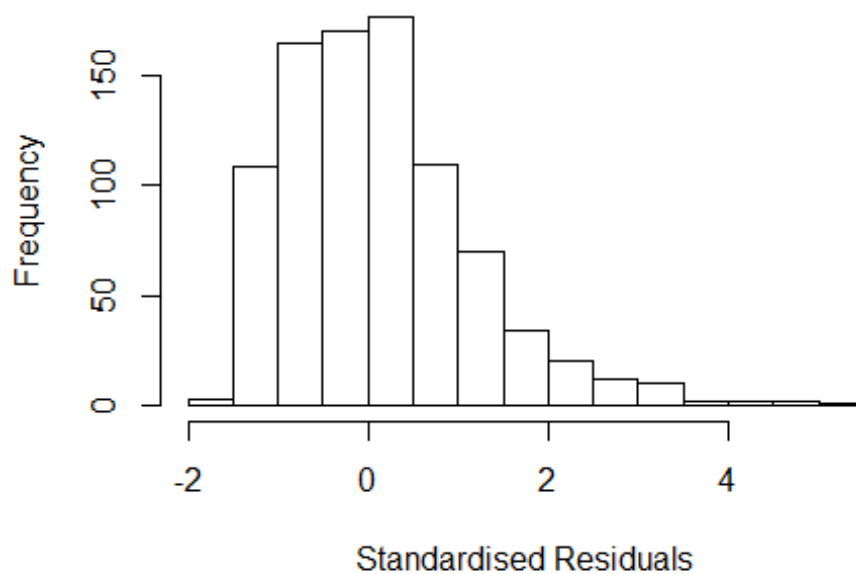


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.9, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.31145891990012"
## [1] "Male first author team size 2018 geometric mean: 1.38016804852938"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3700, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.29956329519794"
## [1] "Male last author team size 2018 geometric mean: 1.39494435218549"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, 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 3.041 1      1.744
## LastAuthorFemale  2.984 1      1.727
## UniqueAuthors    1.653 4      1.065
## Year             1.800 16      1.019
```

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
## 89  60949868527 6.843 2002    1208     2    5.140
## 141 2442613368 4.011 2003    1213     2    3.107
## 143 34247575975 4.335 2003    1213     2    3.257
## 147 70450081408 4.248 2003    1208     2    3.170
## 193 84998075571 3.426 2004    1213     2    2.580
## 206 61449095821 4.199 2005    1213     1    3.179
## 226 43249171214 3.508 2005    1213     2    2.662
## 248 34249137082 3.745 2006    1213     1    2.969
## 252 61049345314 4.319 2006    1213     1    3.368
## 303 61049098694 5.379 2007    1213     1    4.304
## 336 66249150225 4.625 2007    1213     1    3.376
## 360 84998186713 3.843 2007    1213     2    2.594
## 384 61049189557 3.727 2008    1213     1    2.696
## 386 60950625388 3.992 2008    1213     2    2.787
## 439 67649376492 4.345 2008    1213     2    3.140
## 467 84997907764 3.992 2008    1213     2    2.659
## 538 67650816432 4.398 2009    1213     2    3.267
## 635 79952865333 3.540 2010    1213     1    2.659
## 717 84855778105 4.563 2011    1208     2    3.698
## 727 84860870363 4.032 2011    1213     2    3.167
## 839 79960478515 3.818 2011    1208     2    2.779
## 918 84866098127 4.858 2012    1208     2    3.811
## 922 84866119495 4.214 2012    1213     2    2.865
## 930 84865681764 3.712 2012    1213     2    2.839
## 959 84861581198 3.712 2012    1213     2    2.665
```

```

## 960 84861592197 4.875 2012      1213      2      4.002
## 962 84861621527 6.265 2012      1213      2      4.916
## 964 84861632269 4.410 2012      1213      2      3.363
## 965 84861651000 6.164 2012      1213      2      4.815
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7029 -0.7764 -0.0356  0.6663  5.1401
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7134     0.2780   2.57  0.0105 *
## FirstAuthorFemale1 -0.0153     0.1075  -0.14  0.8871
## LastAuthorFemale1  0.1895     0.1073   1.77  0.0778 .
## UniqueAuthors2     0.3019     0.0981   3.08  0.0022 **
## UniqueAuthors3     0.6224     0.1995   3.12  0.0019 **
## UniqueAuthors4     0.7230     0.1786   4.05 5.6e-05 ***
## UniqueAuthors5     0.2939     0.6044   0.49  0.6269
## Year1997          -0.5534     0.3387  -1.63  0.1027
## Year1998           0.1095     0.3140   0.35  0.7274
## Year1999          -0.0604     0.3500  -0.17  0.8630
## Year2000           0.2115     0.3822   0.55  0.5801
## Year2001          -0.0372     0.3057  -0.12  0.9033
## Year2002           0.5134     0.3660   1.40  0.1611
## Year2003           0.1906     0.3405   0.56  0.5758
## Year2004           0.1323     0.3442   0.38  0.7007
## Year2005           0.1327     0.3455   0.38  0.7009
## Year2006           0.0630     0.3346   0.19  0.8507
## Year2007           0.3618     0.2988   1.21  0.2262
## Year2008           0.3174     0.3001   1.06  0.2905
## Year2009           0.1158     0.2959   0.39  0.6957
## Year2010          -0.0068     0.2908  -0.02  0.9813
## Year2011           0.1512     0.2943   0.51  0.6077
## Year2012           0.1598     0.2962   0.54  0.5897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.902
## Multiple R-squared:  0.0508, Adjusted R-squared:  0.0267
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 4 observations c(74,259,795,798)

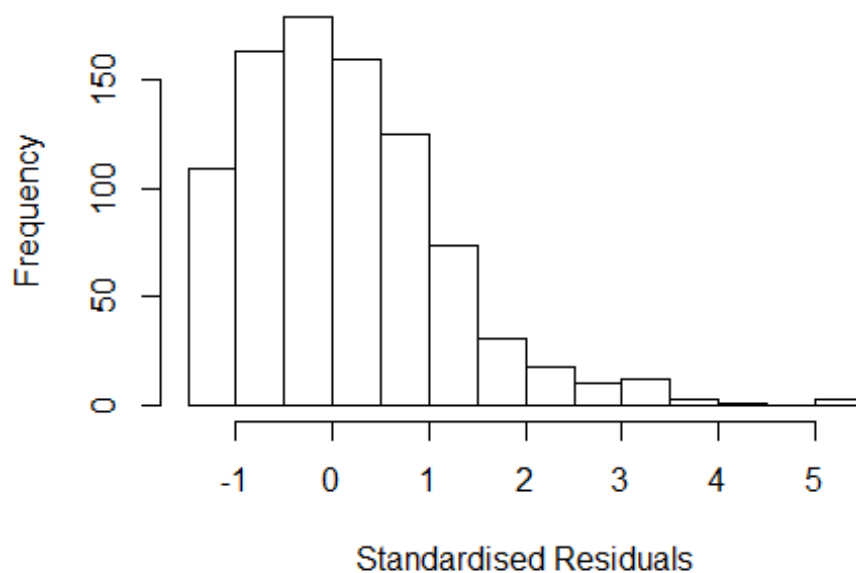
```

```

## are outliers with |weight| = 0 ( < 0.00011);
## 72 weights are ~= 1. The remaining 811 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0106 0.8810 0.9330 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      1.13e-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 3.114 1      1.765
## LastAuthorFemale 3.135 1      1.771
## Year      1.220 16      1.006

```

Residuals from first and last author



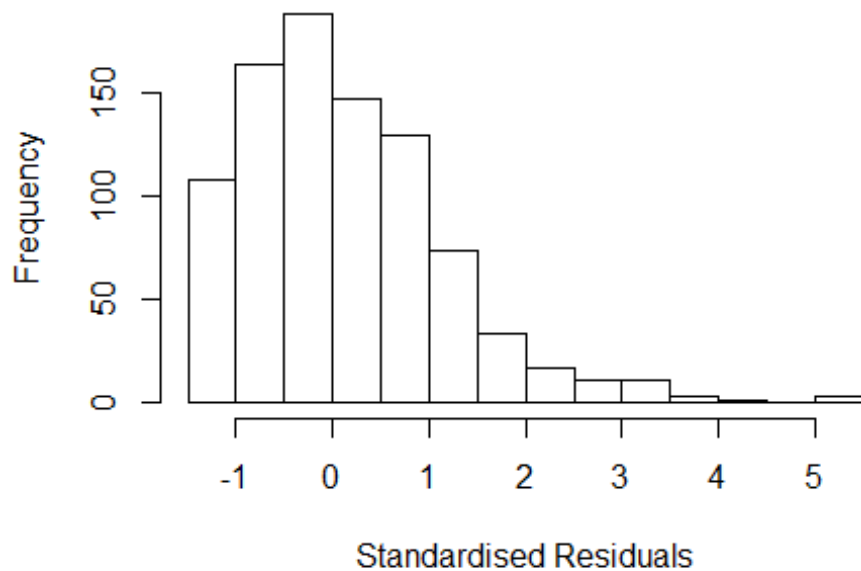

```
## [1] "List of 29 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 89  60949868527 6.843 2002    1208      2    5.416
## 119 33747110111 3.977 2003    1213      2    3.008
## 141  2442613368 4.011 2003    1213      2    3.042
## 143 34247575975 4.335 2003    1213      2    3.191
## 147 70450081408 4.248 2003    1208      2    3.104
## 206 61449095821 4.199 2005    1213      1    3.121
## 226 43249171214 3.508 2005    1213      2    2.605
## 248 34249137082 3.745 2006    1213      1    2.876
## 252 61049345314 4.319 2006    1213      1    3.275
## 303 61049098694 5.379 2007    1213      1    4.281
## 336 66249150225 4.625 2007    1213      1    3.351
## 360 84998186713 3.843 2007    1213      2    2.569
## 384 61049189557 3.727 2008    1213      1    2.642
## 386 60950625388 3.992 2008    1213      2    2.732
## 439 67649376492 4.345 2008    1213      2    3.085
## 467 84997907764 3.992 2008    1213      2    2.907
## 538 67650816432 4.398 2009    1213      2    3.500
## 635 79952865333 3.540 2010    1213      1    2.548
## 717 84855778105 4.563 2011    1208      2    3.610
## 727 84860870363 4.032 2011    1213      2    3.079
## 839 79960478515 3.818 2011    1208      2    2.690
## 918 84866098127 4.858 2012    1208      2    3.730
## 922 84866119495 4.214 2012    1213      2    3.086
## 930 84865681764 3.712 2012    1213      2    2.759
## 959 84861581198 3.712 2012    1213      2    2.584
## 960 84861592197 4.875 2012    1213      2    3.922
## 962 84861621527 6.265 2012    1213      2    5.137
## 964 84861632269 4.410 2012    1213      2    3.282
## 965 84861651000 6.164 2012    1213      2    5.036
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4274 -0.8166 -0.0316  0.6599  5.4156
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.71370    0.27709   2.58    0.01 *
## FirstAuthorFemale1 0.01796    0.11245   0.16    0.87
## LastAuthorFemale1 0.15733    0.11277   1.40    0.16
## Year1997       -0.55178    0.33848  -1.63    0.10
## Year1998        0.14886    0.31439   0.47    0.64
## Year1999        0.00714    0.35307   0.02    0.98
## Year2000        0.23978    0.39930   0.60    0.55
```

```

## Year2001      -0.01216    0.30967   -0.04    0.97
## Year2002      0.53841    0.36680    1.47    0.14
## Year2003      0.25506    0.34049    0.75    0.45
## Year2004      0.21244    0.34151    0.62    0.53
## Year2005      0.18931    0.34257    0.55    0.58
## Year2006      0.15518    0.34045    0.46    0.65
## Year2007      0.38465    0.29819    1.29    0.20
## Year2008      0.37150    0.29910    1.24    0.21
## Year2009      0.18458    0.29565    0.62    0.53
## Year2010      0.10286    0.28916    0.36    0.72
## Year2011      0.23946    0.29380    0.82    0.42
## Year2012      0.23929    0.29444    0.81    0.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.945
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.00638
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(74,795,798) are outliers with |weight| = 0 ( < 0.00011);
## 67 weights are ~= 1. The remaining 817 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0044 0.8810 0.9330 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          1.13e-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 1.139 1          1.067
## Year              1.139 16          1.004

```

Residuals from first author



```
## [1] "List of 29 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 89  60949868527 6.843 2002    1208     2    5.416
## 119 33747110111 3.977 2003    1213     2    3.008
## 141  2442613368 4.011 2003    1213     2    3.042
## 143 34247575975 4.335 2003    1213     2    3.191
## 147 70450081408 4.248 2003    1208     2    3.104
## 206 61449095821 4.199 2005    1213     1    3.121
## 226 43249171214 3.508 2005    1213     2    2.605
## 248 34249137082 3.745 2006    1213     1    2.876
## 252 61049345314 4.319 2006    1213     1    3.275
## 303 61049098694 5.379 2007    1213     1    4.281
## 336 66249150225 4.625 2007    1213     1    3.351
## 360 84998186713 3.843 2007    1213     2    2.569
## 384 61049189557 3.727 2008    1213     1    2.642
## 386 60950625388 3.992 2008    1213     2    2.732
## 439 67649376492 4.345 2008    1213     2    3.085
## 467 84997907764 3.992 2008    1213     2    2.907
## 538 67650816432 4.398 2009    1213     2    3.500
## 635 79952865333 3.540 2010    1213     1    2.548
## 717 84855778105 4.563 2011    1208     2    3.610
## 727 84860870363 4.032 2011    1213     2    3.079
## 839 79960478515 3.818 2011    1208     2    2.690
## 918 84866098127 4.858 2012    1208     2    3.730
## 922 84866119495 4.214 2012    1213     2    3.086
## 930 84865681764 3.712 2012    1213     2    2.759
## 959 84861581198 3.712 2012    1213     2    2.584
```

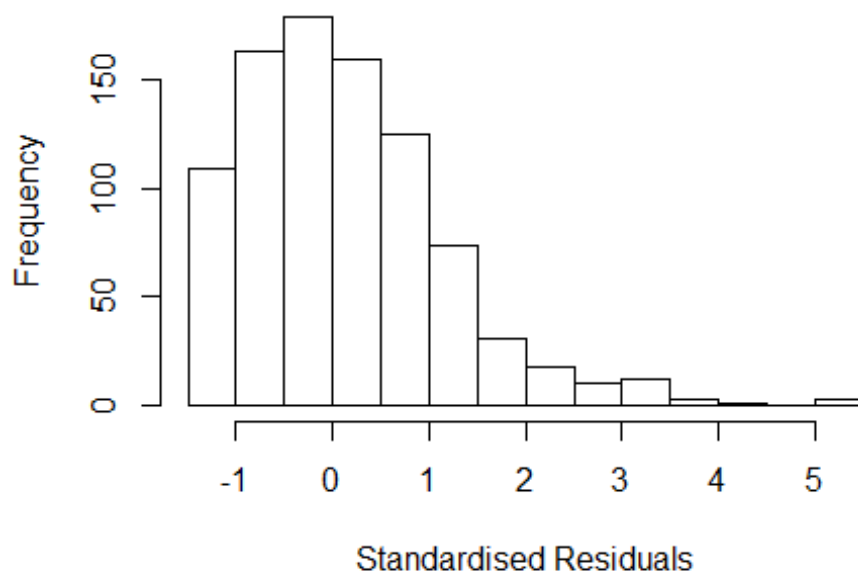
```

## 960 84861592197 4.875 2012      1213      2      3.922
## 962 84861621527 6.265 2012      1213      2      5.137
## 964 84861632269 4.410 2012      1213      2      3.282
## 965 84861651000 6.164 2012      1213      2      5.036
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4106 -0.8292 -0.0272  0.6727  5.4324
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.71372    0.27704     2.58  0.010 *
## FirstAuthorFemale1 0.15179    0.06833     2.22  0.027 *
## Year1997        -0.54490    0.33957    -1.60  0.109
## Year1998         0.14638    0.31495     0.46  0.642
## Year1999         0.01030    0.35386     0.03  0.977
## Year2000         0.24966    0.40032     0.62  0.533
## Year2001         0.00303    0.30886     0.01  0.992
## Year2002         0.54508    0.36388     1.50  0.134
## Year2003         0.27427    0.34129     0.80  0.422
## Year2004         0.21596    0.34149     0.63  0.527
## Year2005         0.20742    0.34285     0.60  0.545
## Year2006         0.18038    0.33984     0.53  0.596
## Year2007         0.40118    0.29793     1.35  0.178
## Year2008         0.37714    0.29909     1.26  0.208
## Year2009         0.19864    0.29495     0.67  0.501
## Year2010         0.11545    0.28907     0.40  0.690
## Year2011         0.24651    0.29406     0.84  0.402
## Year2012         0.24371    0.29441     0.83  0.408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.948
## Multiple R-squared:  0.0248, Adjusted R-squared:  0.00568
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(74,795,798) are outliers with |weight| = 0 ( < 0.00011);
## 66 weights are ~= 1. The remaining 818 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8830 0.9320 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          1.13e-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 1.142 1          1.069
## Year              1.142 16          1.004
```

Residuals from last author



```
## [1] "List of 29 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 89  60949868527 6.843 2002      1208      2      5.416
## 119 33747110111 3.977 2003      1213      2      3.008
## 141  2442613368 4.011 2003      1213      2      3.042
## 143 34247575975 4.335 2003      1213      2      3.191
## 147 70450081408 4.248 2003      1208      2      3.104
## 206 61449095821 4.199 2005      1213      1      3.121
## 226 43249171214 3.508 2005      1213      2      2.605
## 248 34249137082 3.745 2006      1213      1      2.876
```

```

## 252 61049345314 4.319 2006      1213      1      3.275
## 303 61049098694 5.379 2007      1213      1      4.281
## 336 66249150225 4.625 2007      1213      1      3.351
## 360 84998186713 3.843 2007      1213      2      2.569
## 384 61049189557 3.727 2008      1213      1      2.642
## 386 60950625388 3.992 2008      1213      2      2.732
## 439 67649376492 4.345 2008      1213      2      3.085
## 467 84997907764 3.992 2008      1213      2      2.907
## 538 67650816432 4.398 2009      1213      2      3.500
## 635 79952865333 3.540 2010      1213      1      2.548
## 717 84855778105 4.563 2011      1208      2      3.610
## 727 84860870363 4.032 2011      1213      2      3.079
## 839 79960478515 3.818 2011      1208      2      2.690
## 918 84866098127 4.858 2012      1208      2      3.730
## 922 84866119495 4.214 2012      1213      2      3.086
## 930 84865681764 3.712 2012      1213      2      2.759
## 959 84861581198 3.712 2012      1213      2      2.584
## 960 84861592197 4.875 2012      1213      2      3.922
## 962 84861621527 6.265 2012      1213      2      5.137
## 964 84861632269 4.410 2012      1213      2      3.282
## 965 84861651000 6.164 2012      1213      2      5.036
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4262 -0.8182 -0.0332  0.6601  5.4168
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.71374    0.27698   2.58    0.010 *
## LastAuthorFemale1 0.17233    0.06829   2.52    0.012 *
## Year1997       -0.55068    0.33859  -1.63    0.104
## Year1998        0.15104    0.31394   0.48    0.631
## Year1999        0.00788    0.35300   0.02    0.982
## Year2000        0.24171    0.39907   0.61    0.545
## Year2001       -0.00976    0.30918  -0.03    0.975
## Year2002        0.54017    0.36639   1.47    0.141
## Year2003        0.25845    0.34050   0.76    0.448
## Year2004        0.21664    0.34017   0.64    0.524
## Year2005        0.19194    0.34200   0.56    0.575
## Year2006        0.15780    0.34012   0.46    0.643
## Year2007        0.38663    0.29811   1.30    0.195
## Year2008        0.37509    0.29828   1.26    0.209
## Year2009        0.18613    0.29558   0.63    0.529
## Year2010        0.10443    0.28879   0.36    0.718
## Year2011        0.24249    0.29284   0.83    0.408

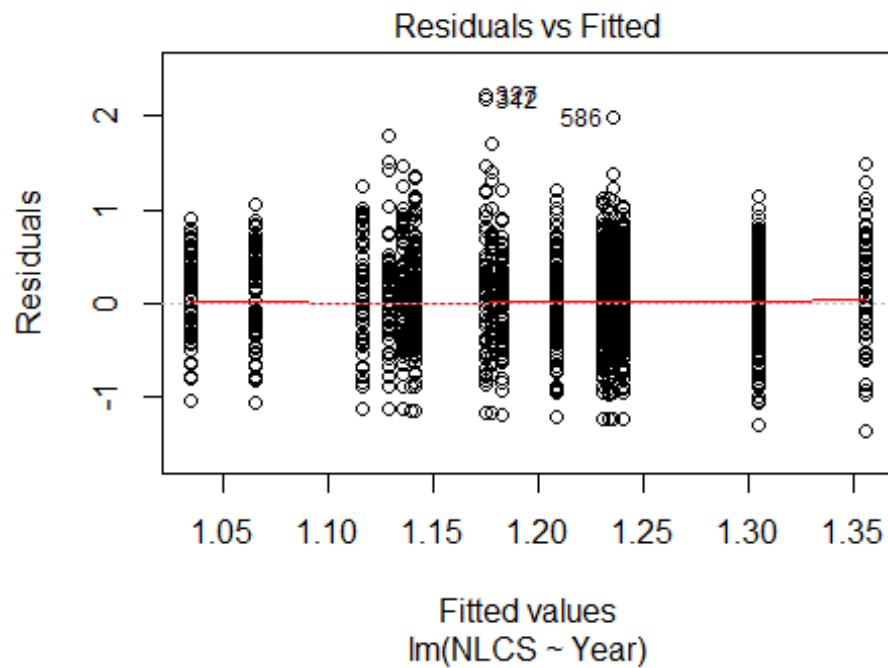
```

```

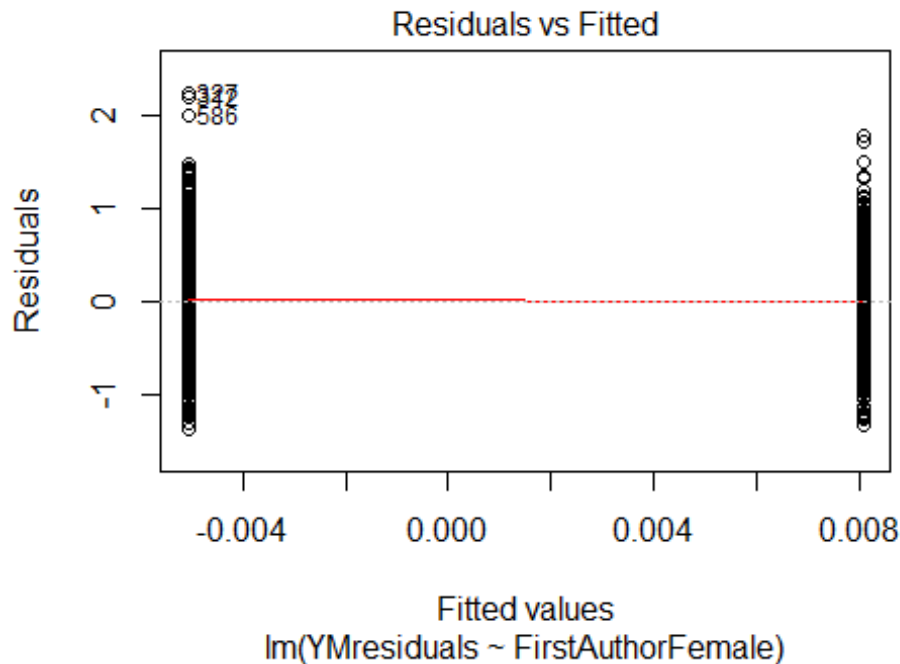
## Year2012          0.24287    0.29368    0.83    0.408
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.951
## Multiple R-squared:  0.0264, Adjusted R-squared:  0.00739
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(74,795,798) are outliers with |weight| = 0 ( < 0.00011);
## 67 weights are ~= 1. The remaining 817 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0061 0.8820 0.9340 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      1.13e-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: 887"
## [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
## 171 167 131 116 180 111 95 131 112 109 137 172 137 204 243
## 2011 2012
## 473 730
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 75 60 68 66 59 39 66 84 74 68 92 133 107 169 183
## 2011 2012
## 361 557
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 52 59 57 56 36 59 68 67 57 85 122 98 143 160

```

```
## 2011 2012
## 318 502
## [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
```

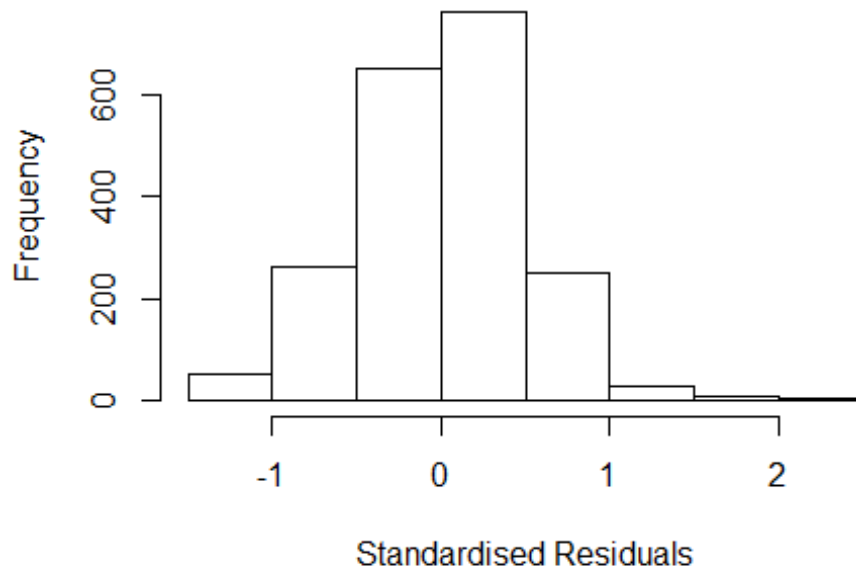


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.3, df = 1, p-value = 0.004
```

```
## [1] "Female first author team size 2018 geometric mean: 4.9656257099333"
## [1] "Male first author team size 2018 geometric mean: 3.8956734768428"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 33000, p-value = 1e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.67291771007089"
## [1] "Male last author team size 2018 geometric mean: 4.29565705529603"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 28000, 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 1.095 1 1.047
## LastAuthorFemale 1.088 1 1.043
## UniqueAuthors 1.349 4 1.038
## Year 1.402 16 1.011
```

Residuals from first and last author and team size



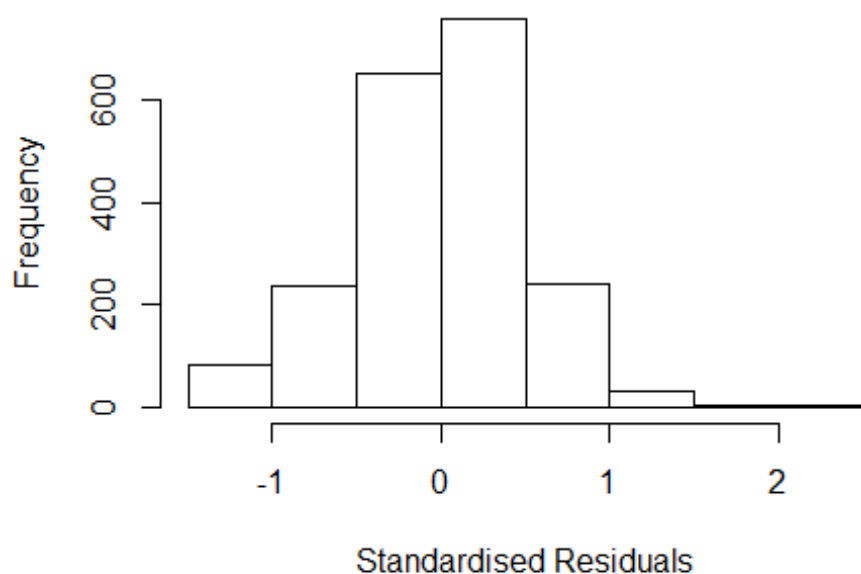
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3852 -0.3294 0.0294 0.3168 2.1365
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1479 0.1055 10.88 < 2e-16 ***
## FirstAuthorFemale1 -0.0214 0.0233 -0.92 0.3589
## LastAuthorFemale1 -0.0612 0.0274 -2.24 0.0255 *
## UniqueAuthors2 0.3189 0.0582 5.47 4.9e-08 ***
## UniqueAuthors3 0.2373 0.0568 4.18 3.0e-05 ***
## UniqueAuthors4 0.3259 0.0595 5.48 4.8e-08 ***
## UniqueAuthors5 0.4040 0.0544 7.43 1.6e-13 ***
## Year1997 -0.2463 0.1603 -1.54 0.1246
## Year1998 -0.3226 0.1138 -2.84 0.0046 **
## Year1999 -0.1547 0.1239 -1.25 0.2121
```

```

## Year2000          -0.2116      0.1324    -1.60    0.1102
## Year2001          -0.2787      0.1262    -2.21    0.0274 *
## Year2002          -0.2941      0.1327    -2.22    0.0268 *
## Year2003          -0.2336      0.1110    -2.11    0.0354 *
## Year2004          -0.3114      0.1160    -2.69    0.0073 **
## Year2005          -0.3244      0.1092    -2.97    0.0030 **
## Year2006          -0.2836      0.1118    -2.54    0.0113 *
## Year2007          -0.2378      0.1064    -2.23    0.0256 *
## Year2008          -0.1816      0.1060    -1.71    0.0868 .
## Year2009          -0.1943      0.1021    -1.90    0.0571 .
## Year2010          -0.1664      0.0986    -1.69    0.0917 .
## Year2011          -0.1265      0.0953    -1.33    0.1845
## Year2012          -0.2126      0.0940    -2.26    0.0238 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.476
## Multiple R-squared:  0.078, Adjusted R-squared:  0.0678
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1863 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.007  0.868  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          4.98e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.073 1 1.036
## LastAuthorFemale 1.090 1 1.044
## Year 1.110 16 1.003

```

Residuals from first and last author



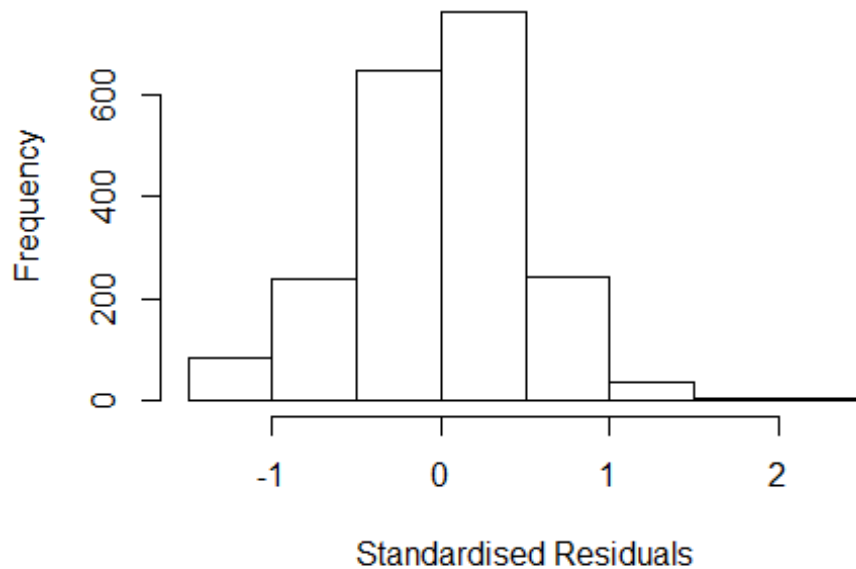
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4145 -0.3300 0.0253 0.3315 2.2711
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.41453 0.09149 15.46 <2e-16 ***
## FirstAuthorFemale1 -0.00146 0.02356 -0.06 0.9506
## LastAuthorFemale1 -0.05475 0.02823 -1.94 0.0526 .
## Year1997 -0.28759 0.15926 -1.81 0.0711 .
## Year1998 -0.34906 0.11345 -3.08 0.0021 **
## Year1999 -0.18420 0.12306 -1.50 0.1346
## Year2000 -0.21115 0.13438 -1.57 0.1163
## Year2001 -0.26513 0.12169 -2.18 0.0295 *
## Year2002 -0.27951 0.13682 -2.04 0.0412 *
## Year2003 -0.22285 0.10876 -2.05 0.0406 *
## Year2004 -0.31353 0.11573 -2.71 0.0068 **
## Year2005 -0.31836 0.11247 -2.83 0.0047 **
```

```

## Year2006          -0.26035    0.11412   -2.28    0.0226 *
## Year2007          -0.23699    0.10798   -2.19    0.0283 *
## Year2008          -0.14429    0.10666   -1.35    0.1763
## Year2009          -0.18041    0.10249   -1.76    0.0785 .
## Year2010          -0.14031    0.09887   -1.42    0.1560
## Year2011          -0.08111    0.09497   -0.85    0.3931
## Year2012          -0.15190    0.09383   -1.62    0.1056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.026, Adjusted R-squared:  0.0172
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 101 is an outlier with |weight| = 0 ( < 5e-05);
## 150 weights are ~= 1. The remaining 1859 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 0.8720 0.9510 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      4.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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
## -1.4067 -0.3218 0.0211 0.3308 2.2779
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.40675 0.09174 15.33 <2e-16 ***
## FirstAuthorFemale1 -0.00944 0.02332 -0.40 0.6857
## Year1997 -0.28668 0.15966 -1.80 0.0727 .
## Year1998 -0.34885 0.11362 -3.07 0.0022 **
## Year1999 -0.18119 0.12296 -1.47 0.1408
## Year2000 -0.21083 0.13551 -1.56 0.1199
## Year2001 -0.26346 0.12210 -2.16 0.0311 *
## Year2002 -0.27725 0.13674 -2.03 0.0427 *
## Year2003 -0.22511 0.10854 -2.07 0.0382 *
## Year2004 -0.31554 0.11566 -2.73 0.0064 **
## Year2005 -0.31811 0.11290 -2.82 0.0049 **
## Year2006 -0.26029 0.11454 -2.27 0.0232 *
```

```

## Year2007          -0.24148      0.10819      -2.23      0.0257 *
## Year2008          -0.14329      0.10689      -1.34      0.1803
## Year2009          -0.18140      0.10277      -1.77      0.0777 .
## Year2010          -0.13985      0.09896      -1.41      0.1578
## Year2011          -0.08477      0.09500      -0.89      0.3723
## Year2012          -0.15449      0.09390      -1.65      0.1001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.024, Adjusted R-squared:  0.0157
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 101 is an outlier with |weight| = 0 ( < 5e-05);
## 153 weights are ~= 1. The remaining 1856 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0004 0.8700 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          4.98e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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
##
## [1] "List of 0 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

```

```

## -1.4143 -0.3301 0.0249 0.3320 2.2714
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4143     0.0915   15.46  <2e-16 ***
## LastAuthorFemale1 -0.0550     0.0279   -1.97   0.0488 *
## Year1997          -0.2877     0.1592   -1.81   0.0708 .
## Year1998          -0.3491     0.1135   -3.08   0.0021 **
## Year1999          -0.1843     0.1229   -1.50   0.1339
## Year2000          -0.2113     0.1343   -1.57   0.1157
## Year2001          -0.2653     0.1217   -2.18   0.0294 *
## Year2002          -0.2796     0.1368   -2.04   0.0412 *
## Year2003          -0.2231     0.1086   -2.05   0.0401 *
## Year2004          -0.3138     0.1155   -2.72   0.0067 **
## Year2005          -0.3185     0.1125   -2.83   0.0047 **
## Year2006          -0.2607     0.1140   -2.29   0.0224 *
## Year2007          -0.2372     0.1079   -2.20   0.0279 *
## Year2008          -0.1446     0.1065   -1.36   0.1747
## Year2009          -0.1807     0.1024   -1.76   0.0778 .
## Year2010          -0.1406     0.0986   -1.43   0.1540
## Year2011          -0.0814     0.0947   -0.86   0.3903
## Year2012          -0.1523     0.0934   -1.63   0.1033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.026, Adjusted R-squared:  0.0177
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 101 is an outlier with |weight| = 0 ( < 5e-05);
## 148 weights are ~ = 1. The remaining 1861 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 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           4.98e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)

```



```

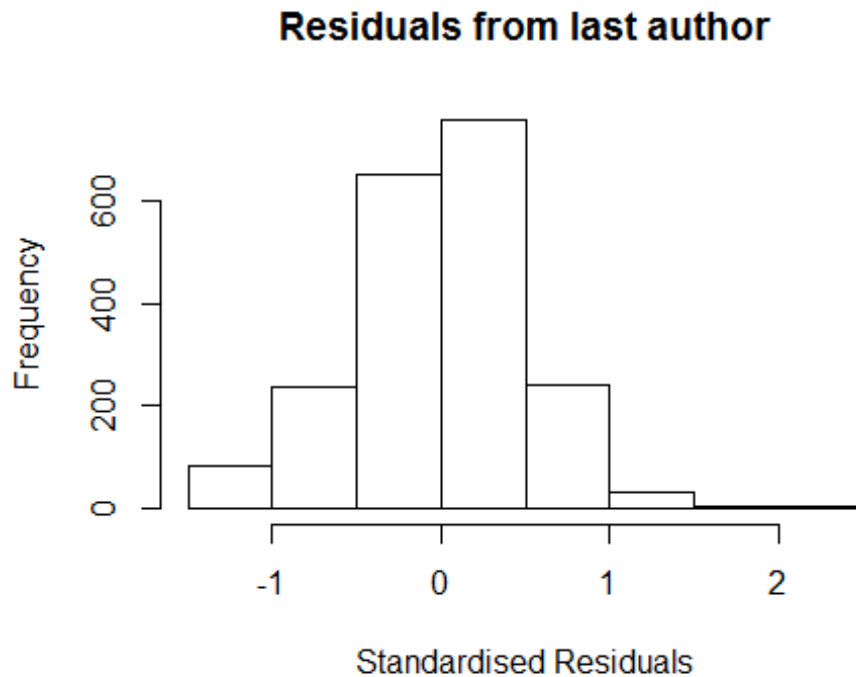
## [1] "Sample size for the above analysis: 2010"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1301"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2007 2008 2009 2011 2012
##    2    3    3    5    3
##
## 2007 2008 2009 2011 2012
##    2    1    1    5    3
##
## 2007 2008 2009 2011 2012
##    2    1    1    4    2
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.47213595499958"
## [1] "Male first author team size 2018 geometric mean: 5.1924941018511"

## 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 = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5"
## [1] "Male last author team size 2018 geometric mean: 4.86459855819557"

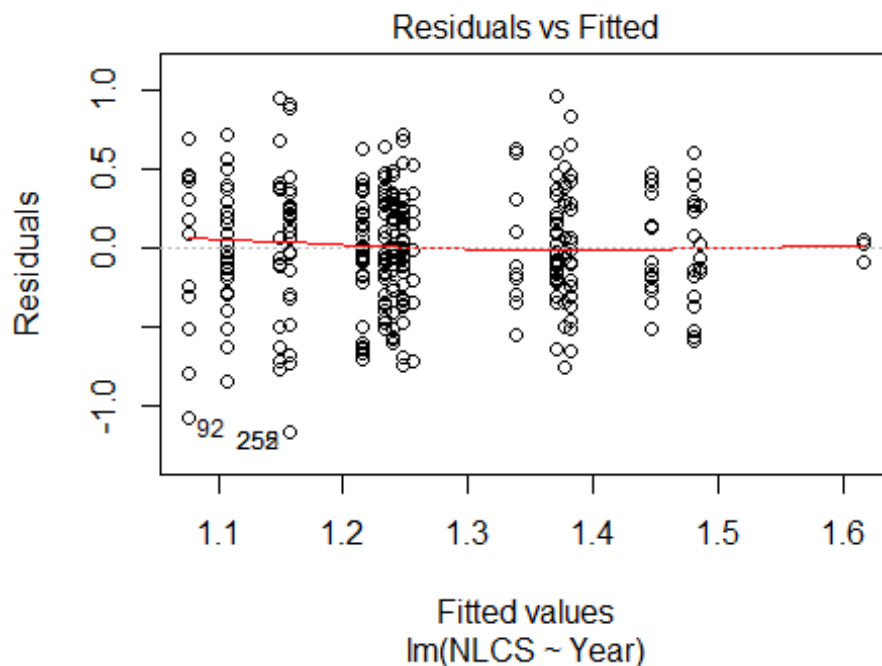
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2.5, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## [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: 10"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1302"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
## [1] "gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    22    7   10   14   16   24   20   24   16   21   32   33   41   38   38
## 2011 2012
##    34   46
##
```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 3 6 8 10 13 13 19 12 14 29 29 33 29 29
## 2011 2012
## 27 33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 2 6 6 7 11 13 19 12 13 23 26 29 24 27
## 2011 2012
## 24 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 = 28, df = 16, p-value = 0.03
```

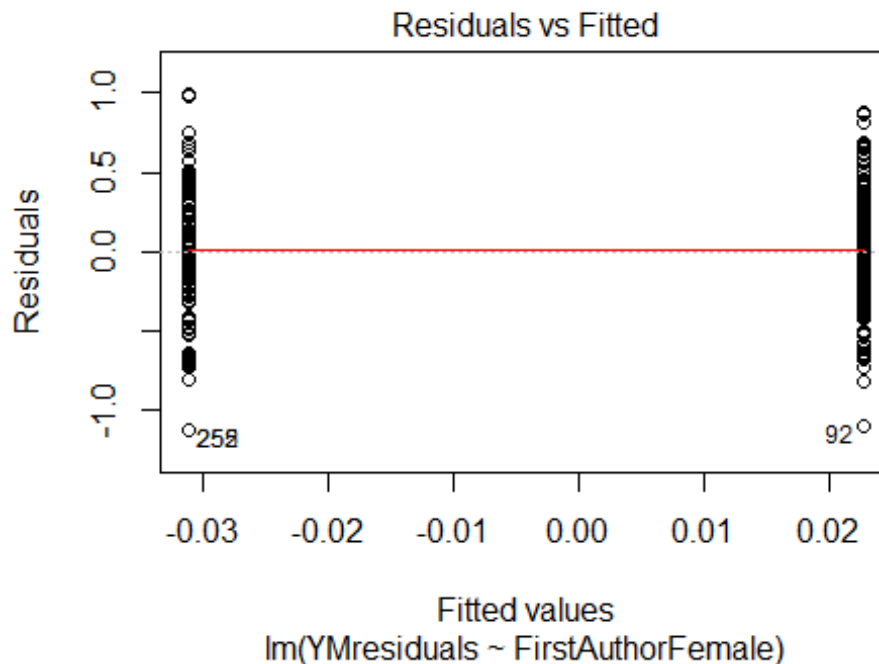


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
##
## [1] "Female first author team size 2018 geometric mean: 3.7362477871571"
## [1] "Male first author team size 2018 geometric mean: 6.12382623932545"
```

```
## 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 = 0.008
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.42361190440849"
## [1] "Male last author team size 2018 geometric mean: 5.38193315741092"

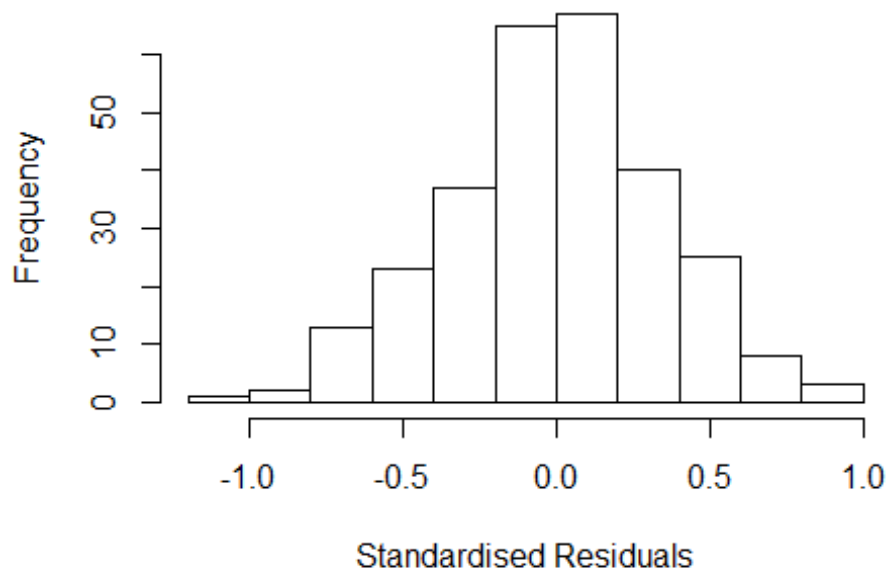
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 86, 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 1.891 1          1.375
```

## LastAuthorFemale	2.270	1	1.507
## UniqueAuthors	8.085	4	1.299
## Year	9.475	16	1.073

Residuals from first and last author and team size



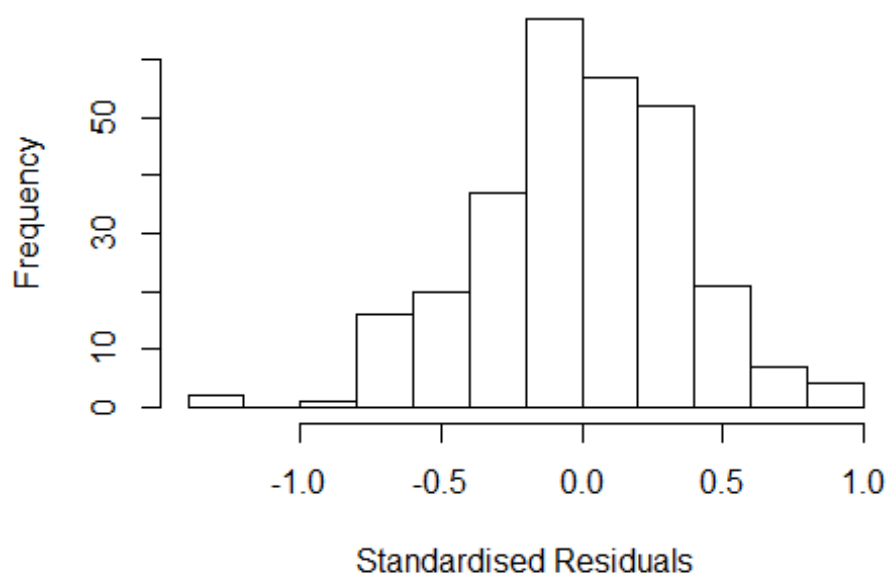
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13046 -0.22564 0.00539 0.22702 0.96013
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19940 0.09930 12.08 < 2e-16 ***
## FirstAuthorFemale1 -0.00193 0.05189 -0.04 0.97033
## LastAuthorFemale1 -0.00772 0.05368 -0.14 0.88572
## UniqueAuthors2 0.33676 0.09937 3.39 0.00081 ***
## UniqueAuthors3 0.28707 0.09315 3.08 0.00228 **
## UniqueAuthors4 0.14763 0.10260 1.44 0.15138
```

```

## UniqueAuthors5      0.28487    0.09481    3.00  0.00292 **
## Year1997             0.11168    0.10755    1.04  0.30005
## Year1998             0.08567    0.11168    0.77  0.44369
## Year1999            -0.34289    0.18334   -1.87  0.06257 .
## Year2000            -0.06094    0.15990   -0.38  0.70344
## Year2001            -0.24259    0.23594   -1.03  0.30483
## Year2002            -0.32387    0.20830   -1.55  0.12120
## Year2003            -0.05677    0.12972   -0.44  0.66200
## Year2004            -0.06065    0.14180   -0.43  0.66922
## Year2005            -0.02796    0.13103   -0.21  0.83117
## Year2006            -0.22758    0.11809   -1.93  0.05505 .
## Year2007            -0.20886    0.11838   -1.76  0.07885 .
## Year2008            -0.24317    0.11770   -2.07  0.03981 *
## Year2009            -0.20126    0.11836   -1.70  0.09025 .
## Year2010            -0.09421    0.10813   -0.87  0.38438
## Year2011            -0.32798    0.11248   -2.92  0.00385 **
## Year2012            -0.20911    0.11422   -1.83  0.06829 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.148, Adjusted R-squared:  0.0761
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 258 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.210  0.854  0.951  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      3.52e-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 1.584 1          1.259
## LastAuthorFemale  1.750 1          1.323
## Year              2.304 16          1.026

```

Residuals from first and last author



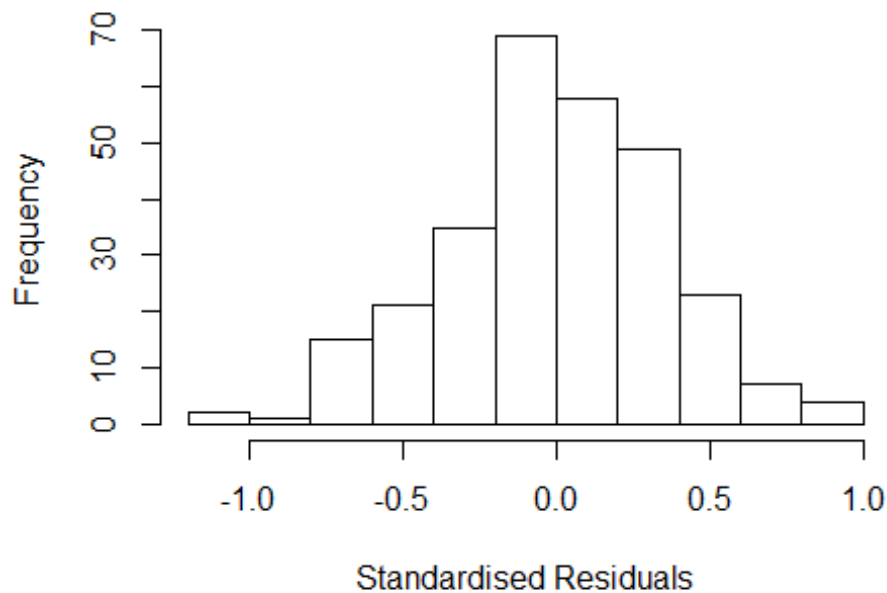
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20548 -0.21836 -0.00377 0.26522 0.98883
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.41152 0.11116 12.70 <2e-16 ***
## FirstAuthorFemale1 0.00726 0.04949 0.15 0.883
## LastAuthorFemale1 0.01138 0.04978 0.23 0.819
## Year1997 0.15834 0.12146 1.30 0.193
## Year1998 0.06403 0.12315 0.52 0.604
## Year1999 -0.28306 0.18753 -1.51 0.132
## Year2000 -0.02292 0.18899 -0.12 0.904
## Year2001 -0.22468 0.22238 -1.01 0.313
## Year2002 -0.27792 0.23459 -1.18 0.237
## Year2003 -0.06095 0.14638 -0.42 0.677
## Year2004 -0.01905 0.16812 -0.11 0.910
## Year2005 0.03427 0.14696 0.23 0.816
```

```

## Year2006      -0.20918    0.13303   -1.57    0.117
## Year2007      -0.21946    0.13399   -1.64    0.103
## Year2008      -0.20770    0.12962   -1.60    0.110
## Year2009      -0.19485    0.13134   -1.48    0.139
## Year2010      -0.06935    0.12420   -0.56    0.577
## Year2011      -0.28407    0.12614   -2.25    0.025 *
## Year2012      -0.18830    0.12758   -1.48    0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0799, Adjusted R-squared:  0.0174
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 268 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.866  0.947  0.904  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      3.52e-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 1.51 1      1.229
## Year              1.51 16      1.013

```


Residuals from first author



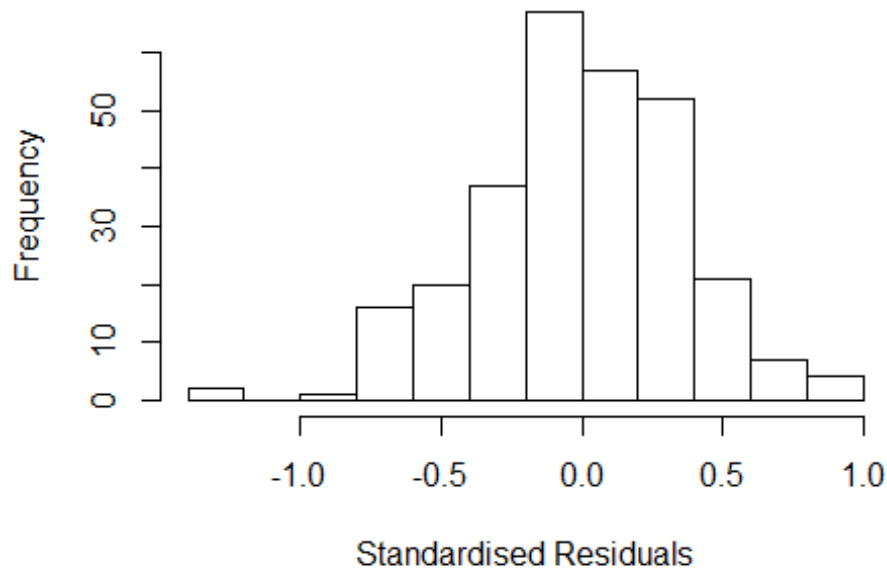
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19803 -0.21539 -0.00429  0.26076  0.98660
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4169    0.1089   13.01  <2e-16 ***
## FirstAuthorFemale1  0.0101    0.0484    0.21   0.834
## Year1997          0.1615    0.1212    1.33   0.184
## Year1998          0.0631    0.1235    0.51   0.610
## Year1999         -0.2812    0.1875   -1.50   0.135
## Year2000         -0.0268    0.1883   -0.14   0.887
## Year2001         -0.2290    0.2190   -1.05   0.297
## Year2002         -0.2836    0.2296   -1.24   0.218
## Year2003         -0.0637    0.1472   -0.43   0.666
## Year2004         -0.0235    0.1667   -0.14   0.888
## Year2005          0.0330    0.1474    0.22   0.823
## Year2006         -0.2108    0.1333   -1.58   0.115
```

```

## Year2007          -0.2233      0.1336   -1.67    0.096 .
## Year2008          -0.2102      0.1304   -1.61    0.108
## Year2009          -0.1979      0.1308   -1.51    0.131
## Year2010          -0.0725      0.1238   -0.59    0.558
## Year2011          -0.2890      0.1254   -2.30    0.022 *
## Year2012          -0.1920      0.1271   -1.51    0.132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.0799, Adjusted R-squared:  0.0211
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 264 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.213  0.865  0.946  0.902  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      3.52e-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 1.663 1      1.289
## Year      1.663 16      1.016

```

Residuals from last author



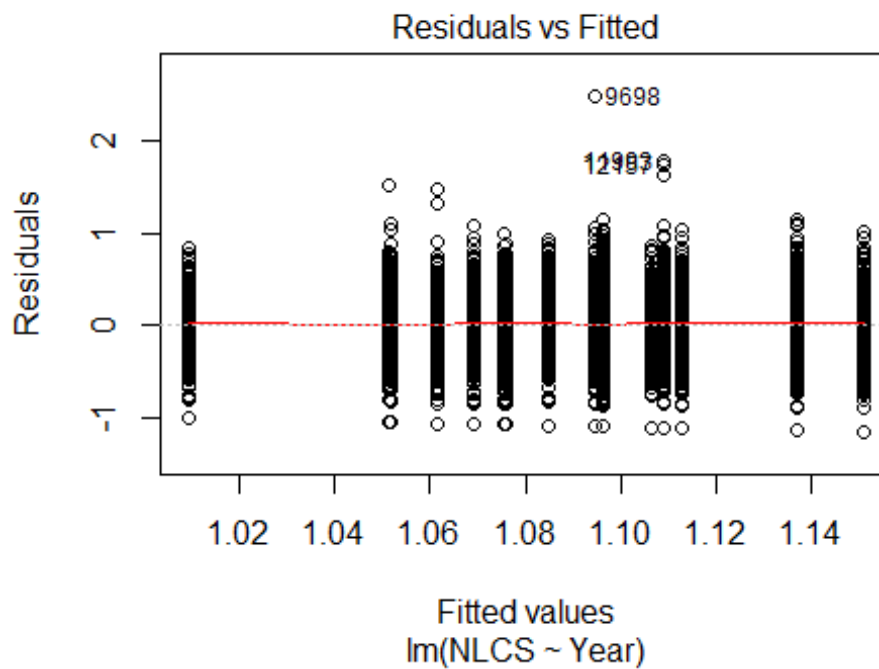
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20856 -0.22281 -0.00551  0.26819  0.98520
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4134     0.1110   12.74  <2e-16 ***
## LastAuthorFemale1  0.0131     0.0486    0.27   0.788
## Year1997          0.1621     0.1167    1.39   0.166
## Year1998          0.0638     0.1227    0.52   0.603
## Year1999         -0.2821     0.1876   -1.50   0.134
## Year2000         -0.0217     0.1887   -0.11   0.909
## Year2001         -0.2270     0.2207   -1.03   0.305
## Year2002         -0.2766     0.2314   -1.20   0.233
## Year2003         -0.0592     0.1457   -0.41   0.685
## Year2004         -0.0162     0.1646   -0.10   0.922
## Year2005          0.0345     0.1463    0.24   0.814
## Year2006         -0.2071     0.1303   -1.59   0.113
```

```

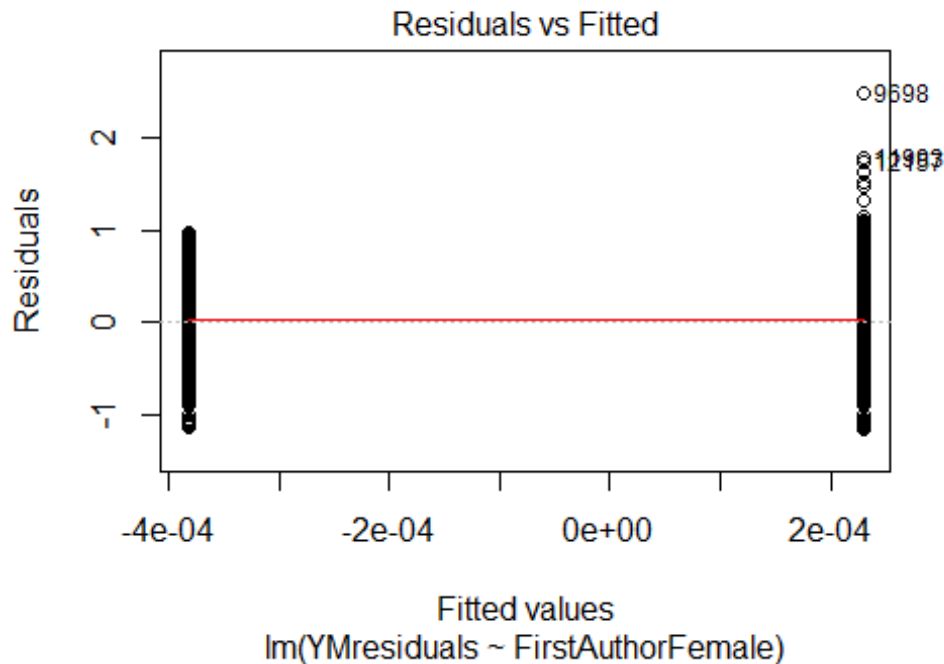
## Year2007          -0.2179      0.1323    -1.65      0.101
## Year2008          -0.2063      0.1287    -1.60      0.110
## Year2009          -0.1919      0.1276    -1.50      0.134
## Year2010          -0.0676      0.1233    -0.55      0.584
## Year2011          -0.2811      0.1238    -2.27      0.024 *
## Year2012          -0.1858      0.1256    -1.48      0.140
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0798, Adjusted R-squared:  0.021
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 266 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.209  0.867  0.946  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      3.52e-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: 284"
## [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
##  649  601  568  567  583  542  544  506  505  565  595  608  602  660  722
## 2011 2012
##  707  699
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  398  366  382  386  320  276  385  338  364  432  435  438  435  479  506
## 2011 2012

```

```
## 503 492
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 353 325 346 349 288 252 330 303 327 388 403 396 402 423 459
## 2011 2012
## 445 442
## [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
```

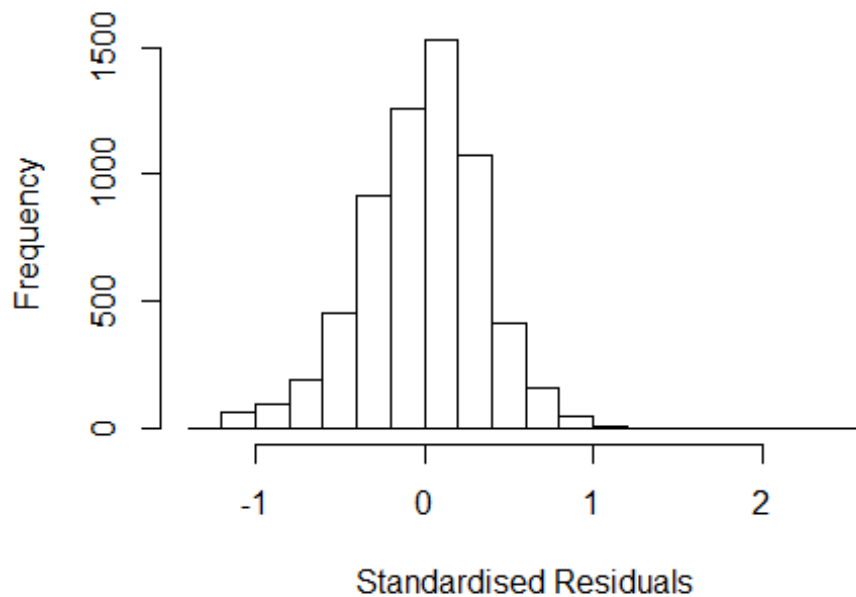


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 18, df = 1, p-value = 2e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 5.32339341519321"
## [1] "Male first author team size 2018 geometric mean: 4.65012014190225"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.05777795263837"
## [1] "Male last author team size 2018 geometric mean: 4.8699983880474"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7800, 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 1.023 1      1.011
## LastAuthorFemale  1.029 1      1.014
## UniqueAuthors    1.098 4      1.012
## Year              1.127 16     1.004
```

Residuals from first and last author and team size



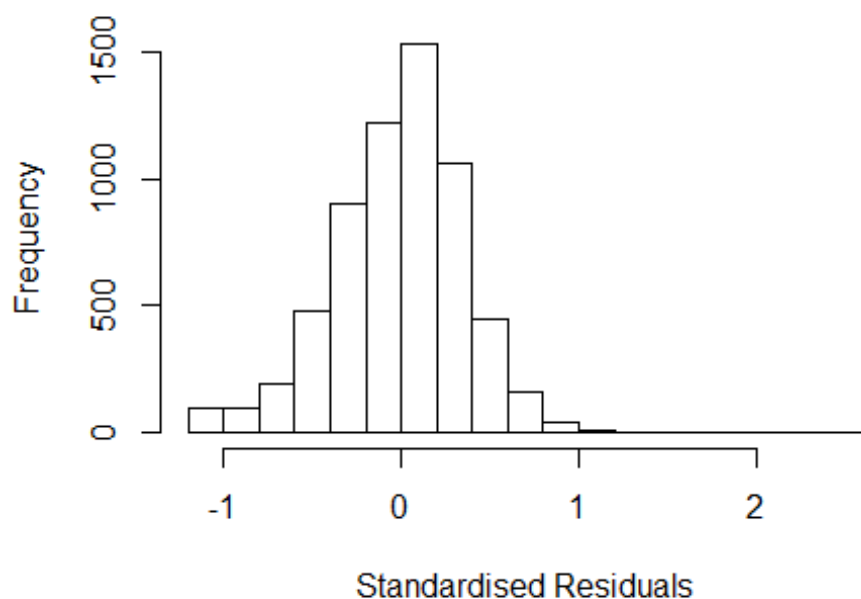
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 9698 75249087100 3.574 2009      1303      6      2.529
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.267 -0.232  0.017  0.224  2.529
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94484    0.04192   22.54 < 2e-16 ***
## FirstAuthorFemale1 -0.01082    0.00912   -1.19  0.23526
## LastAuthorFemale1 -0.00420    0.01330   -0.32  0.75213
## UniqueAuthors2     0.18467    0.03768    4.90 9.7e-07 ***
## UniqueAuthors3     0.20607    0.03702    5.57 2.7e-08 ***
## UniqueAuthors4     0.22866    0.03705    6.17 7.2e-10 ***
## UniqueAuthors5     0.31153    0.03653    8.53 < 2e-16 ***
## Year1997          0.01071    0.03148    0.34 0.73364
## Year1998         -0.00412    0.03203   -0.13 0.89765
## Year1999         -0.06868    0.02842   -2.42 0.01568 *
```

```

## Year2000      -0.07144      0.02998      -2.38      0.01719      *
## Year2001      -0.13833      0.03216      -4.30      1.7e-05      ***
## Year2002      -0.10398      0.02860      -3.64      0.00028      ***
## Year2003      -0.06455      0.02892      -2.23      0.02564      *
## Year2004      -0.10702      0.02848      -3.76      0.00017      ***
## Year2005      -0.13224      0.02701      -4.90      1.0e-06      ***
## Year2006      -0.12512      0.02799      -4.47      7.9e-06      ***
## Year2007      -0.09885      0.02798      -3.53      0.00041      ***
## Year2008      -0.07652      0.02766      -2.77      0.00568      **
## Year2009      -0.10558      0.02784      -3.79      0.00015      ***
## Year2010      -0.10100      0.02746      -3.68      0.00024      ***
## Year2011      -0.08598      0.02828      -3.04      0.00238      **
## Year2012      -0.09160      0.02862      -3.20      0.00138      **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.333
## Multiple R-squared:  0.0486, Adjusted R-squared:  0.0452
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4510,5587,5588,5658)
## are outliers with |weight| = 0 ( < 1.6e-05);
## 515 weights are ~ = 1. The remaining 5712 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8710 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           1.60e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.014 1 1.007
## LastAuthorFemale 1.022 1 1.011
## Year 1.034 16 1.001

```


Residuals from first and last author



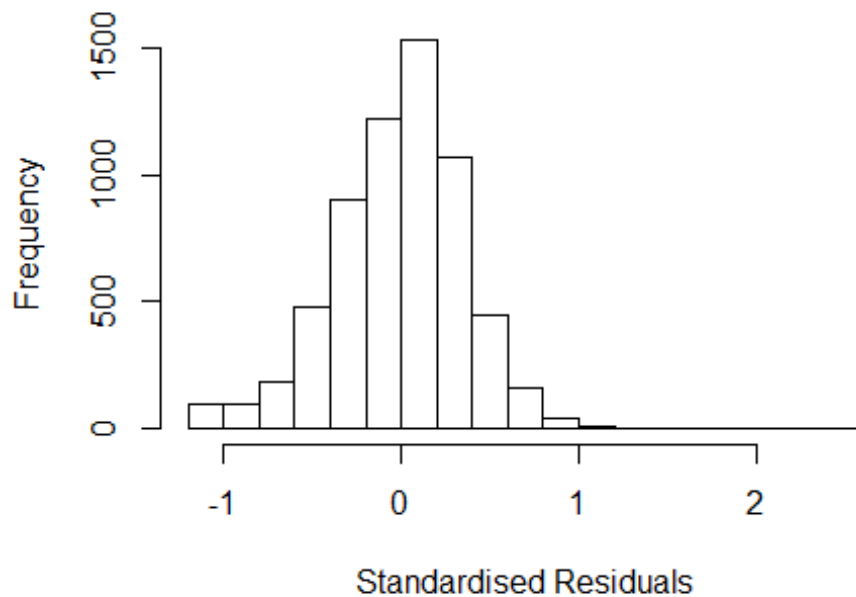
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1745 -0.2386 0.0205 0.2231 2.4854
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.159717 0.022190 52.26 < 2e-16 ***
## FirstAuthorFemale1 0.000292 0.009262 0.03 0.97481
## LastAuthorFemale1 -0.002221 0.013592 -0.16 0.87021
## Year1997 0.014781 0.031382 0.47 0.63764
## Year1998 -0.004246 0.032140 -0.13 0.89489
## Year1999 -0.057584 0.028241 -2.04 0.04149 *
## Year2000 -0.056939 0.030137 -1.89 0.05889 .
## Year2001 -0.115825 0.032199 -3.60 0.00032 ***
## Year2002 -0.082948 0.028655 -2.89 0.00381 **
## Year2003 -0.048634 0.029196 -1.67 0.09580 .
## Year2004 -0.087490 0.028137 -3.11 0.00188 **
## Year2005 -0.103885 0.026975 -3.85 0.00012 ***
```

```

## Year2006          -0.103836    0.027899    -3.72    0.00020 ***
## Year2007          -0.070477    0.027895    -2.53    0.01155 *
## Year2008          -0.048781    0.027566    -1.77    0.07684 .
## Year2009          -0.071099    0.027739    -2.56    0.01040 *
## Year2010          -0.067361    0.027312    -2.47    0.01368 *
## Year2011          -0.051456    0.028176    -1.83    0.06787 .
## Year2012          -0.060503    0.028596    -2.12    0.03440 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.00941,    Adjusted R-squared:  0.00654
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4510,5587,5588,5658)
## are outliers with |weight| = 0 ( < 1.6e-05);
## 512 weights are ~ = 1. The remaining 5715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0094 0.8690 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          1.60e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



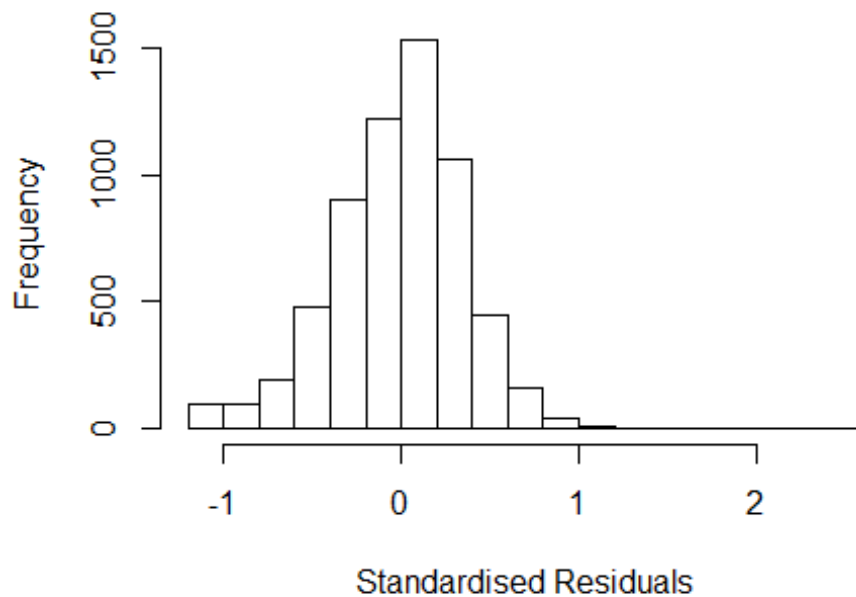
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1743 -0.2391 0.0205 0.2231 2.4857
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.159553 0.022200 52.23 < 2e-16 ***
## FirstAuthorFemale1 0.000232 0.009258 0.03 0.98004
## Year1997 0.014700 0.031378 0.47 0.63945
## Year1998 -0.004269 0.032132 -0.13 0.89431
## Year1999 -0.057678 0.028223 -2.04 0.04103 *
## Year2000 -0.056994 0.030136 -1.89 0.05864 .
## Year2001 -0.115905 0.032175 -3.60 0.00032 ***
## Year2002 -0.083143 0.028612 -2.91 0.00368 **
## Year2003 -0.048726 0.029182 -1.67 0.09503 .
## Year2004 -0.087599 0.028121 -3.12 0.00185 **
## Year2005 -0.104009 0.026950 -3.86 0.00011 ***
## Year2006 -0.104024 0.027847 -3.74 0.00019 ***
```

```

## Year2007          -0.070586    0.027875    -2.53    0.01136 *
## Year2008          -0.048900    0.027535    -1.78    0.07580 .
## Year2009          -0.071301    0.027661    -2.58    0.00997 **
## Year2010          -0.067573    0.027228    -2.48    0.01310 *
## Year2011          -0.051663    0.028105    -1.84    0.06608 .
## Year2012          -0.060728    0.028554    -2.13    0.03348 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.00941,    Adjusted R-squared:  0.0067
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4510,5587,5588,5658)
## are outliers with |weight| = 0 ( < 1.6e-05);
## 517 weights are ~ = 1. The remaining 5710 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0093 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          1.60e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



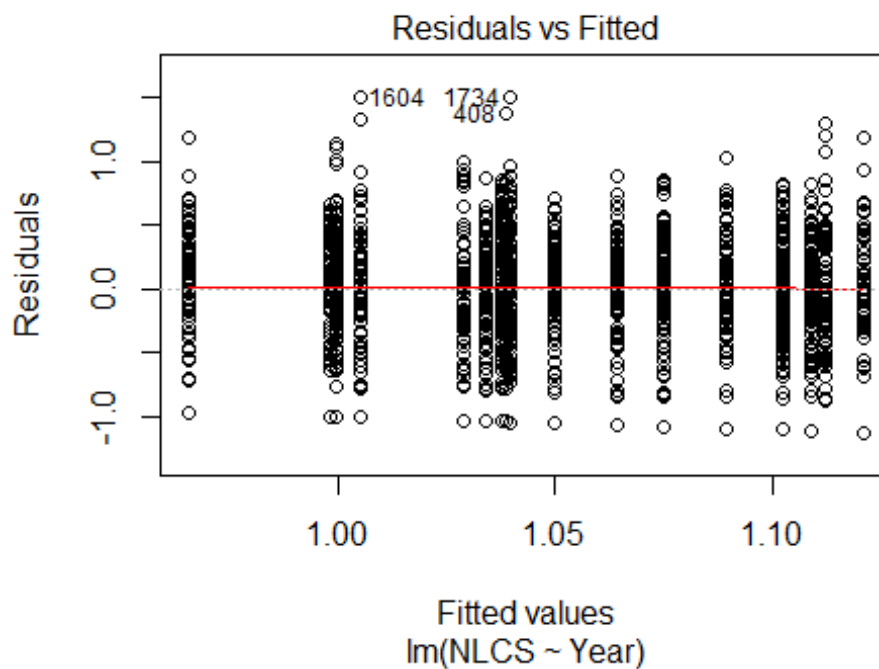
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1746 -0.2386 0.0206 0.2231 2.4853
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15981 0.02192 52.90 < 2e-16 ***
## LastAuthorFemale1 -0.00221 0.01359 -0.16 0.87100
## Year1997 0.01478 0.03138 0.47 0.63766
## Year1998 -0.00425 0.03214 -0.13 0.89484
## Year1999 -0.05758 0.02824 -2.04 0.04150 *
## Year2000 -0.05693 0.03014 -1.89 0.05899 .
## Year2001 -0.11581 0.03221 -3.60 0.00033 ***
## Year2002 -0.08293 0.02865 -2.89 0.00382 **
## Year2003 -0.04862 0.02920 -1.67 0.09591 .
## Year2004 -0.08745 0.02812 -3.11 0.00188 **
## Year2005 -0.10387 0.02697 -3.85 0.00012 ***
## Year2006 -0.10382 0.02790 -3.72 0.00020 ***
```

```

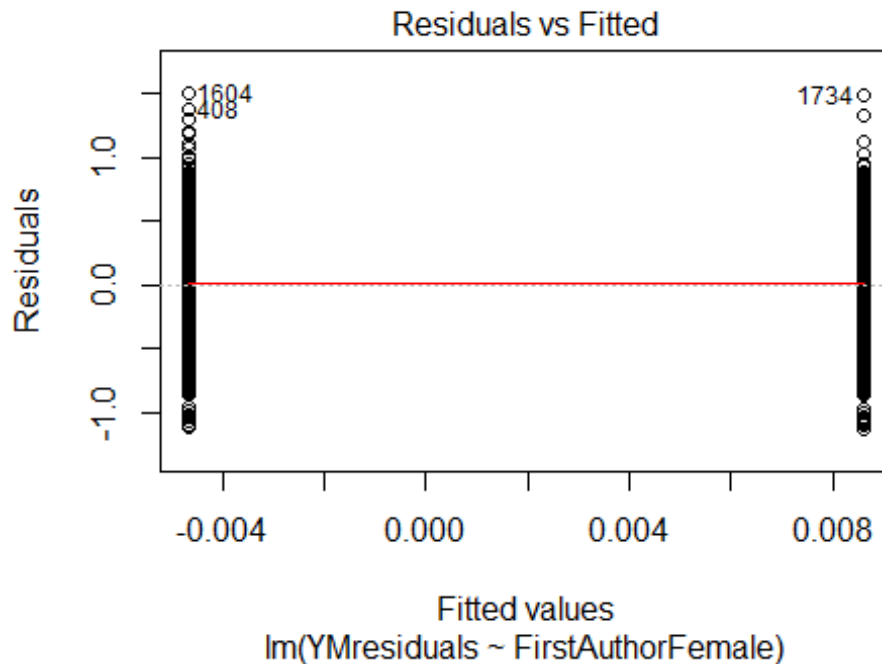
## Year2007          -0.07046      0.02790    -2.53  0.01157 *
## Year2008          -0.04876      0.02756    -1.77  0.07691 .
## Year2009          -0.07108      0.02774    -2.56  0.01043 *
## Year2010          -0.06734      0.02731    -2.47  0.01371 *
## Year2011          -0.05143      0.02817    -1.83  0.06794 .
## Year2012          -0.06048      0.02858    -2.12  0.03437 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.00941,    Adjusted R-squared:  0.0067
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(4510,5587,5588,5658)
## are outliers with |weight| = 0 ( < 1.6e-05);
## 512 weights are ~ = 1. The remaining 5715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0094 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      1.60e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6231"
## [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
## 171 175 176 164 160 142 138 140 134 186 162 152 169 182 206
## 2011 2012
## 201 200
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 95 94 77 88 72 65 80 77 71 120 112 86 111 115 132
## 2011 2012
## 126 131
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 86 86 68 82 65 60 65 65 57 102 104 67 85 103 115
## 2011 2012
## 110 113
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```

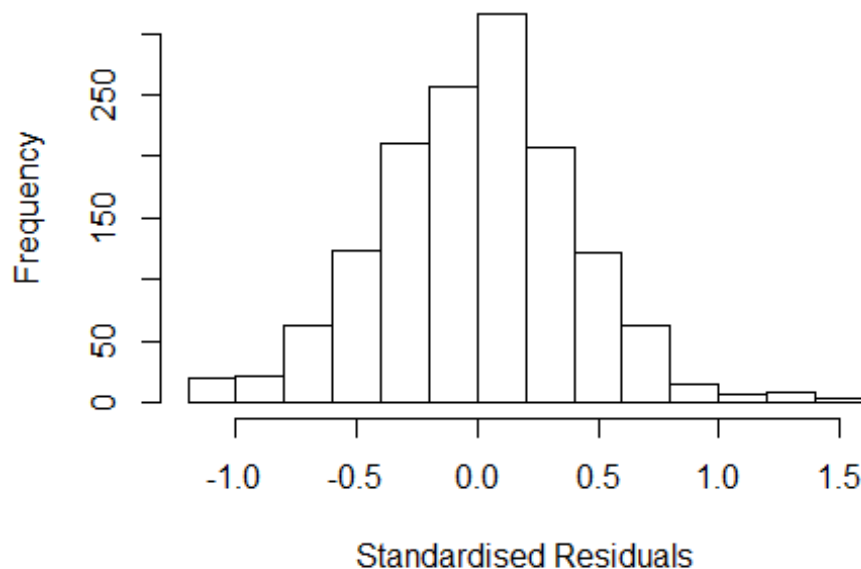


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.6e-06, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.93133690958885"
## [1] "Male first author team size 2018 geometric mean: 3.6821000931795"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.7106847728945"
## [1] "Male last author team size 2018 geometric mean: 3.80669588041042"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, 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 1.122 1      1.059
## LastAuthorFemale  1.072 1      1.035
## UniqueAuthors    1.413 4      1.044
## Year              1.540 16     1.014
```


Residuals from first and last author and team size



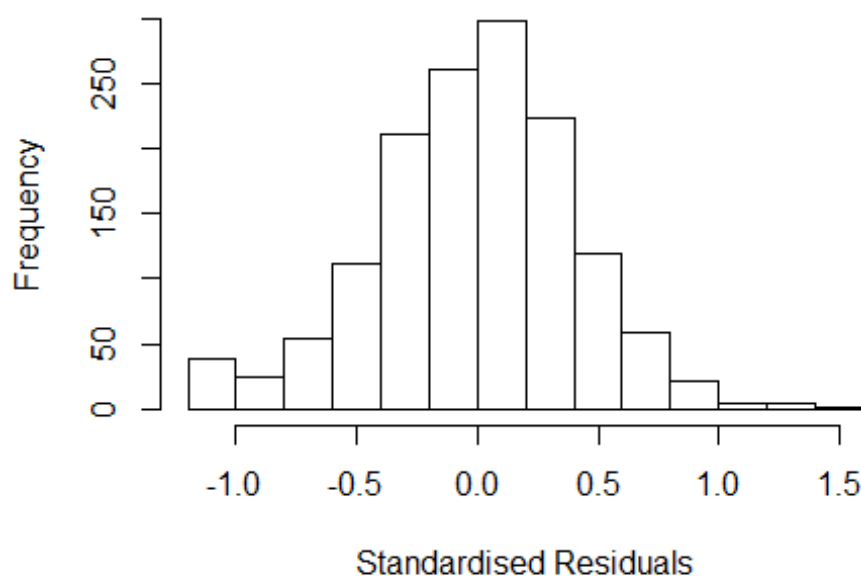
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1531 -0.2649  0.0195  0.2540  1.4928
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.71946    0.08037   8.95 < 2e-16 ***
## FirstAuthorFemale1 -0.00898    0.02331  -0.39  0.7001
## LastAuthorFemale1  0.00764    0.03105   0.25  0.8056
## UniqueAuthors2    0.24890    0.06871   3.62  0.0003 ***
## UniqueAuthors3    0.32014    0.06724   4.76  2.1e-06 ***
## UniqueAuthors4    0.35046    0.06687   5.24  1.8e-07 ***
## UniqueAuthors5    0.34622    0.06638   5.22  2.1e-07 ***
## Year1997          0.05986    0.06530   0.92  0.3594
## Year1998          0.09236    0.06475   1.43  0.1539
## Year1999          0.05556    0.06436   0.86  0.3881
```

```

## Year2000      -0.03576    0.06784   -0.53    0.5983
## Year2001      0.08880    0.06662    1.33    0.1828
## Year2002      0.02903    0.06751    0.43    0.6672
## Year2003      0.04830    0.06718    0.72    0.4722
## Year2004      0.01933    0.07225    0.27    0.7891
## Year2005      0.00955    0.06463    0.15    0.8826
## Year2006      0.00852    0.06543    0.13    0.8964
## Year2007      0.09552    0.07253    1.32    0.1881
## Year2008      0.07071    0.05898    1.20    0.2307
## Year2009      0.08534    0.05837    1.46    0.1439
## Year2010      0.07270    0.06108    1.19    0.2341
## Year2011      0.00649    0.06545    0.10    0.9210
## Year2012      0.07874    0.05827    1.35    0.1768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.0595, Adjusted R-squared:  0.0448
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 115 weights are ~= 1. The remaining 1318 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.103  0.864  0.952  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      6.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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  1.053 1      1.026
## Year              1.133 16      1.004

```

Residuals from first and last author



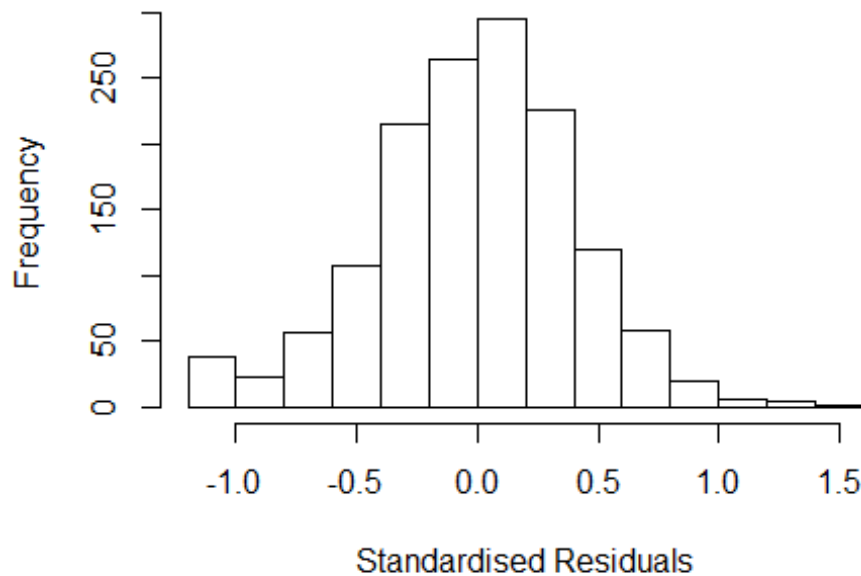
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1458 -0.2581 0.0104 0.2556 1.5023
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00630 0.04728 21.29 <2e-16 ***
## FirstAuthorFemale1 0.01424 0.02339 0.61 0.543
## LastAuthorFemale1 0.01476 0.03225 0.46 0.647
## Year1997 0.05425 0.06685 0.81 0.417
## Year1998 0.08259 0.06408 1.29 0.198
## Year1999 0.06883 0.06319 1.09 0.276
## Year2000 -0.02830 0.06870 -0.41 0.680
## Year2001 0.09807 0.06592 1.49 0.137
## Year2002 0.01743 0.06849 0.25 0.799
## Year2003 0.04594 0.06647 0.69 0.490
## Year2004 -0.00456 0.07342 -0.06 0.950
## Year2005 0.01255 0.06379 0.20 0.844
```

```

## Year2006      -0.00292    0.06495   -0.04    0.964
## Year2007      0.11048    0.07233    1.53    0.127
## Year2008      0.08829    0.05936    1.49    0.137
## Year2009      0.10696    0.05963    1.79    0.073 .
## Year2010      0.09852    0.06165    1.60    0.110
## Year2011      0.02169    0.06471    0.34    0.738
## Year2012      0.10054    0.05856    1.72    0.086 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.000348
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.864  0.951  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      6.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



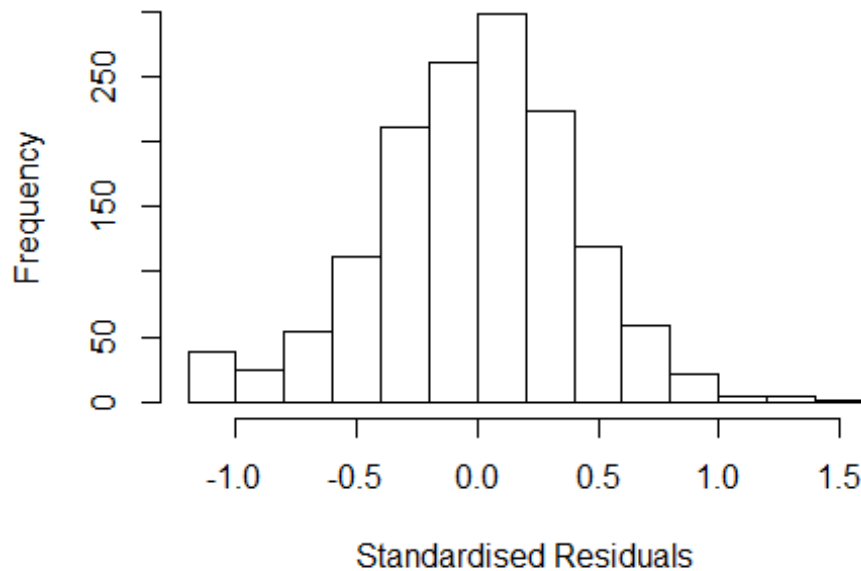
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1347 -0.2564 0.0119 0.2598 1.5004
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00824 0.04712 21.40 <2e-16 ***
## FirstAuthorFemale1 0.01550 0.02339 0.66 0.507
## Year1997 0.05383 0.06683 0.81 0.421
## Year1998 0.08228 0.06406 1.28 0.199
## Year1999 0.06744 0.06298 1.07 0.284
## Year2000 -0.02958 0.06852 -0.43 0.666
## Year2001 0.09748 0.06598 1.48 0.140
## Year2002 0.01766 0.06853 0.26 0.797
## Year2003 0.04458 0.06646 0.67 0.502
## Year2004 -0.00465 0.07365 -0.06 0.950
## Year2005 0.01334 0.06372 0.21 0.834
## Year2006 -0.00230 0.06496 -0.04 0.972
```

```

## Year2007          0.11093      0.07256      1.53      0.127
## Year2008          0.08856      0.05941      1.49      0.136
## Year2009          0.10712      0.05962      1.80      0.073 .
## Year2010          0.09913      0.06161      1.61      0.108
## Year2011          0.02240      0.06474      0.35      0.729
## Year2012          0.10089      0.05856      1.72      0.085 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.0127, Adjusted R-squared:  0.000854
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1315 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.107  0.865  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      6.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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      1.002

```

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
## -1.1377 -0.2578 0.0116 0.2593 1.5106
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01013 0.04644 21.75 <2e-16 ***
## LastAuthorFemale1 0.01682 0.03222 0.52 0.602
## Year1997 0.05445 0.06681 0.81 0.415
## Year1998 0.08301 0.06424 1.29 0.196
## Year1999 0.07059 0.06337 1.11 0.266
## Year2000 -0.02658 0.06863 -0.39 0.699
## Year2001 0.09986 0.06593 1.51 0.130
## Year2002 0.01926 0.06837 0.28 0.778
## Year2003 0.04747 0.06660 0.71 0.476
## Year2004 -0.00371 0.07367 -0.05 0.960
## Year2005 0.01227 0.06376 0.19 0.847
## Year2006 -0.00217 0.06506 -0.03 0.973
```

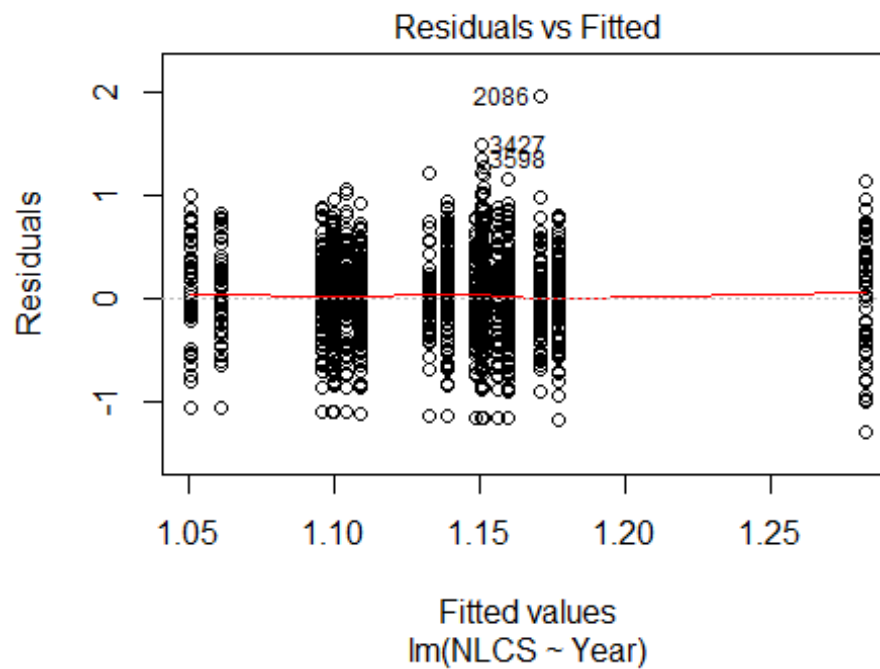
```

## Year2007      0.11072      0.07209      1.54      0.125
## Year2008      0.08864      0.05933      1.49      0.135
## Year2009      0.10816      0.05978      1.81      0.071 .
## Year2010      0.09867      0.06167      1.60      0.110
## Year2011      0.02313      0.06466      0.36      0.721
## Year2012      0.10390      0.05825      1.78      0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.000709
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1309 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.865  0.951  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      6.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1433"
## [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
## 152 158 147 140 145 155 167 161 148 181 185 201 235 259 223
## 2011 2012
## 255 267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 64 77 59 54 55 90 93 79 109 97 107 132 158 116
## 2011 2012

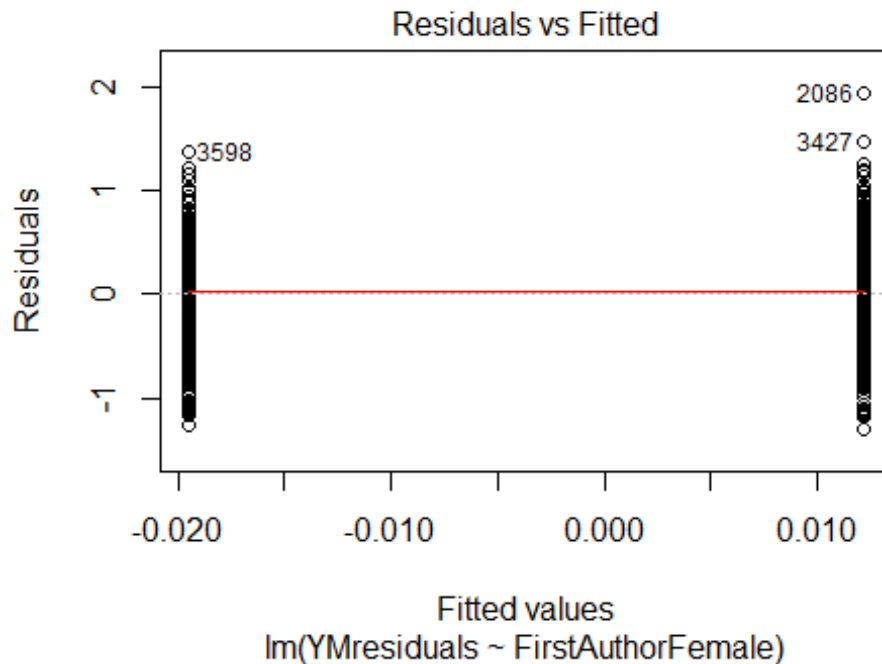
```



```
## 171 153
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 58 70 52 48 46 78 79 64 94 87 93 118 132 106
## 2011 2012
## 155 138
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 73, df = 16, p-value = 4e-09
```

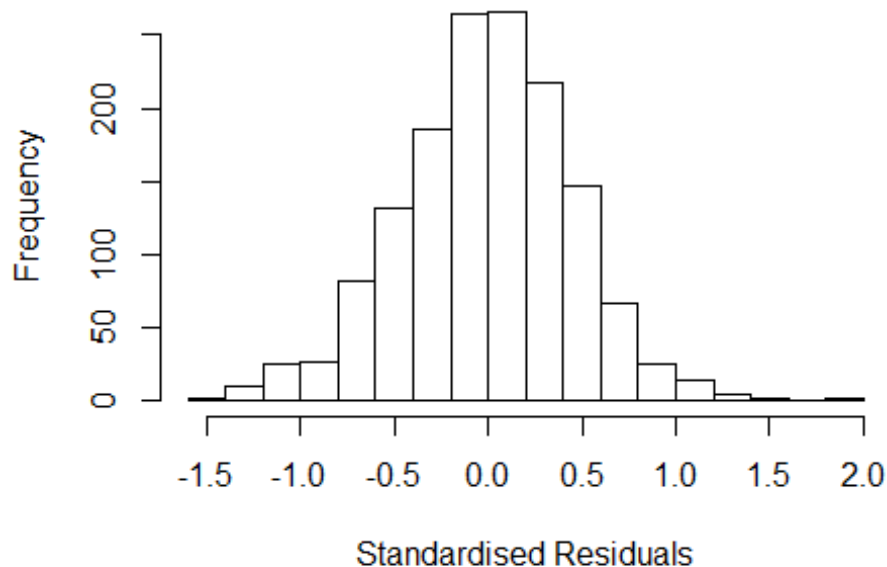


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.048, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 4.49870410603167"
## [1] "Male first author team size 2018 geometric mean: 4.34226516616552"
##
## 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] "Female last author team size 2018 geometric mean: 4.40855548305777"
## [1] "Male last author team size 2018 geometric mean: 4.39530353849173"
##
## 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 1.080 1          1.039
## LastAuthorFemale  1.062 1          1.031
## UniqueAuthors    1.342 4          1.037
## Year              1.422 16         1.011
```

Residuals from first and last author and team size



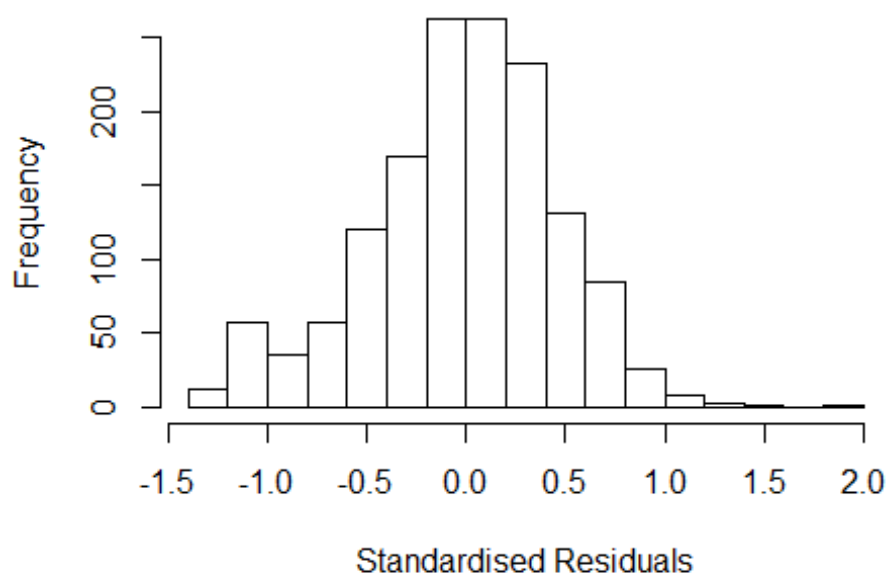
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.45486 -0.28439 0.00657 0.28096 1.85555
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.85836 0.13230 6.49 1.2e-10 ***
## FirstAuthorFemale1 -0.04551 0.02525 -1.80 0.072 .
## LastAuthorFemale1 -0.00252 0.03060 -0.08 0.934
## UniqueAuthors2 0.48118 0.06659 7.23 8.0e-13 ***
## UniqueAuthors3 0.51338 0.06362 8.07 1.5e-15 ***
## UniqueAuthors4 0.57726 0.06441 8.96 < 2e-16 ***
## UniqueAuthors5 0.64201 0.06256 10.26 < 2e-16 ***
## Year1997 -0.14706 0.14215 -1.03 0.301
## Year1998 -0.12227 0.12954 -0.94 0.345
## Year1999 -0.24470 0.13093 -1.87 0.062 .
```

```

## Year2000      -0.24365    0.13915   -1.75    0.080 .
## Year2001      -0.17820    0.12552   -1.42    0.156
## Year2002      -0.22649    0.12584   -1.80    0.072 .
## Year2003      -0.26169    0.12444   -2.10    0.036 *
## Year2004      -0.25709    0.13191   -1.95    0.051 .
## Year2005      -0.19659    0.12246   -1.61    0.109
## Year2006      -0.23392    0.12280   -1.90    0.057 .
## Year2007      -0.27498    0.12144   -2.26    0.024 *
## Year2008      -0.29421    0.12217   -2.41    0.016 *
## Year2009      -0.22900    0.12317   -1.86    0.063 .
## Year2010      -0.23667    0.12216   -1.94    0.053 .
## Year2011      -0.21162    0.12209   -1.73    0.083 .
## Year2012      -0.20223    0.12392   -1.63    0.103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.133, Adjusted R-squared:  0.12
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1349 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0222 0.8650 0.9510 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      6.82e-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 1.045 1      1.022
## LastAuthorFemale  1.037 1      1.019
## Year              1.074 16      1.002

```

Residuals from first and last author



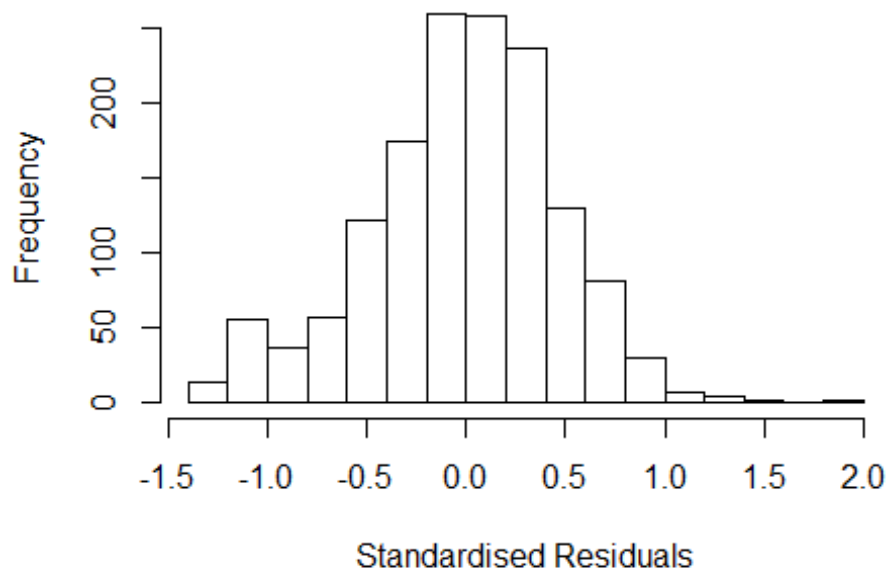
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3282 -0.2846  0.0116  0.2956  1.9622
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3282     0.1086   12.24  <2e-16 ***
## FirstAuthorFemale1 -0.0233     0.0261   -0.89   0.373
## LastAuthorFemale1  -0.0233     0.0325   -0.72   0.473
## Year1997          -0.1214     0.1398   -0.87   0.385
## Year1998          -0.1413     0.1244   -1.14   0.256
## Year1999          -0.2361     0.1252   -1.89   0.060 .
## Year2000          -0.2215     0.1365   -1.62   0.105
## Year2001          -0.2070     0.1211   -1.71   0.088 .
## Year2002          -0.2048     0.1201   -1.71   0.088 .
## Year2003          -0.1979     0.1174   -1.69   0.092 .
## Year2004          -0.2083     0.1254   -1.66   0.097 .
## Year2005          -0.1356     0.1165   -1.16   0.245
```

```

## Year2006          -0.1684      0.1159   -1.45    0.146
## Year2007          -0.1849      0.1148   -1.61    0.108
## Year2008          -0.2042      0.1160   -1.76    0.079 .
## Year2009          -0.1563      0.1163   -1.34    0.179
## Year2010          -0.1909      0.1167   -1.64    0.102
## Year2011          -0.1465      0.1168   -1.25    0.210
## Year2012          -0.1143      0.1191   -0.96    0.337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.0108, Adjusted R-squared:  -0.00147
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1336 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.010  0.860  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      6.82e-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 1.041 1      1.020
## Year              1.041 16      1.001

```

Residuals from first author



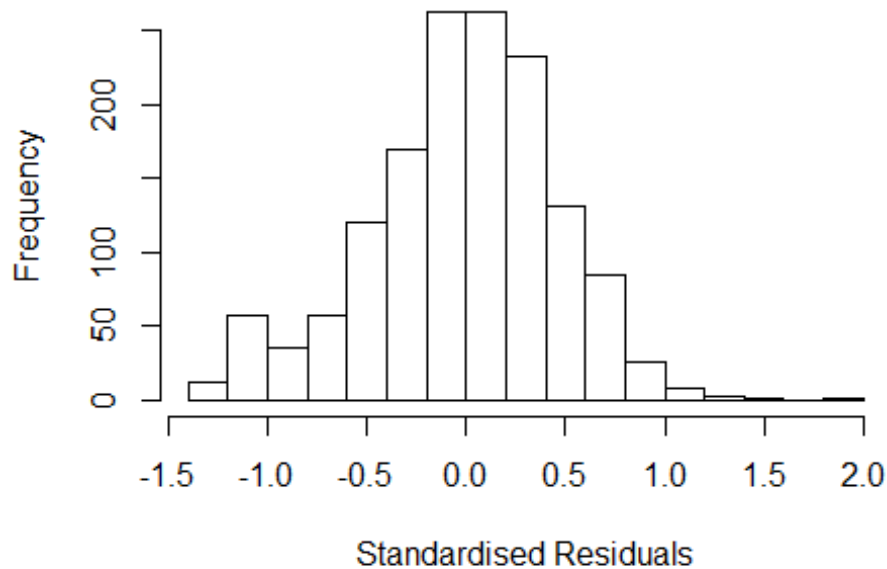
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3243 -0.2851  0.0116  0.2970  1.9644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3243     0.1075   12.32  <2e-16 ***
## FirstAuthorFemale1 -0.0256     0.0261   -0.98   0.327
## Year1997         -0.1231     0.1397   -0.88   0.378
## Year1998         -0.1414     0.1237   -1.14   0.253
## Year1999         -0.2374     0.1246   -1.90   0.057 .
## Year2000         -0.2219     0.1357   -1.63   0.102
## Year2001         -0.2048     0.1200   -1.71   0.088 .
## Year2002         -0.2043     0.1194   -1.71   0.087 .
## Year2003         -0.1973     0.1166   -1.69   0.091 .
## Year2004         -0.2082     0.1245   -1.67   0.095 .
## Year2005         -0.1351     0.1156   -1.17   0.243
## Year2006         -0.1667     0.1148   -1.45   0.147
```

```

## Year2007          -0.1856      0.1137   -1.63    0.103
## Year2008          -0.2045      0.1152   -1.78    0.076 .
## Year2009          -0.1561      0.1155   -1.35    0.177
## Year2010          -0.1909      0.1161   -1.64    0.100
## Year2011          -0.1475      0.1162   -1.27    0.205
## Year2012          -0.1161      0.1190   -0.98    0.330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0104, Adjusted R-squared:  -0.00125
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 132 weights are ~= 1. The remaining 1335 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0116 0.8600 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      6.82e-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 1.033 1          1.016
## Year            1.033 16          1.001

```


Residuals from last author



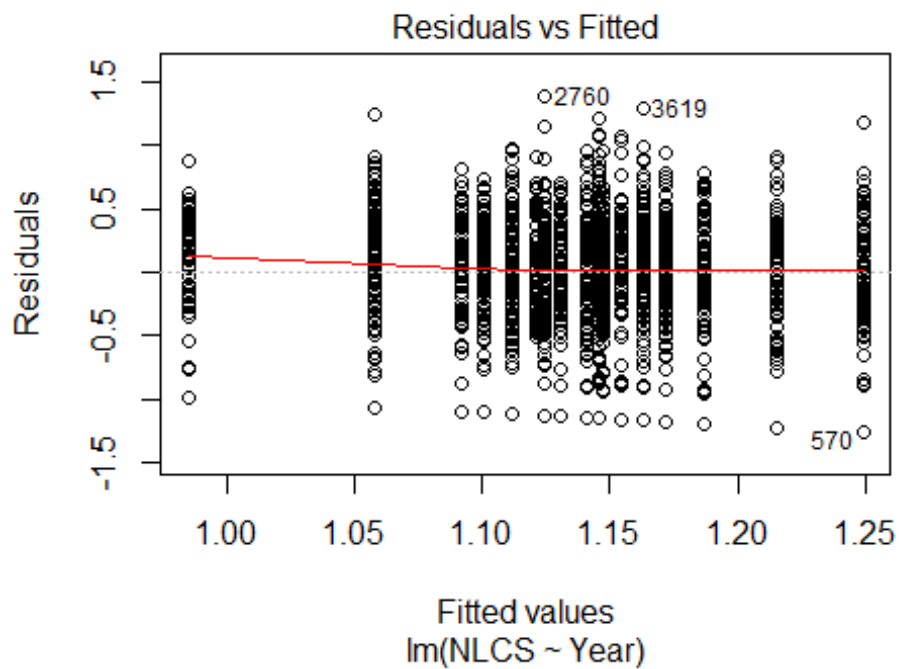
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3233 -0.2878 0.0122 0.2986 1.9715
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3233 0.1075 12.31 <2e-16 ***
## LastAuthorFemale1 -0.0267 0.0326 -0.82 0.412
## Year1997 -0.1221 0.1392 -0.88 0.381
## Year1998 -0.1450 0.1236 -1.17 0.241
## Year1999 -0.2398 0.1243 -1.93 0.054 .
## Year2000 -0.2263 0.1355 -1.67 0.095 .
## Year2001 -0.2080 0.1201 -1.73 0.083 .
## Year2002 -0.2078 0.1193 -1.74 0.082 .
## Year2003 -0.2047 0.1161 -1.76 0.078 .
## Year2004 -0.2100 0.1241 -1.69 0.091 .
## Year2005 -0.1371 0.1156 -1.19 0.236
## Year2006 -0.1728 0.1145 -1.51 0.132
```

```

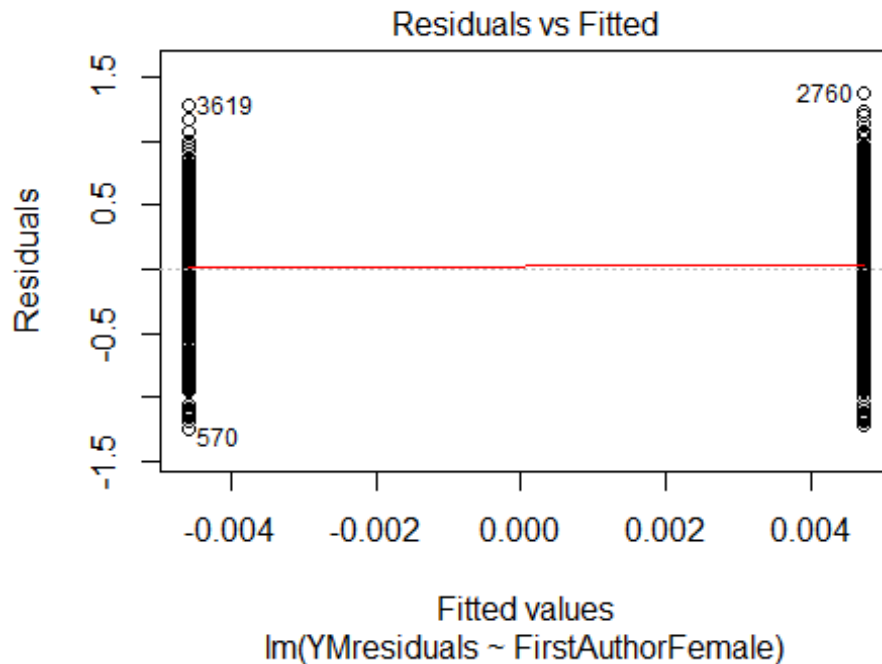
## Year2007          -0.1880      0.1136    -1.66      0.098 .
## Year2008          -0.2093      0.1148    -1.82      0.068 .
## Year2009          -0.1603      0.1153    -1.39      0.165
## Year2010          -0.1952      0.1157    -1.69      0.092 .
## Year2011          -0.1491      0.1161    -1.28      0.199
## Year2012          -0.1211      0.1186    -1.02      0.307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0102, Adjusted R-squared:  -0.00141
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1343 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.010  0.863  0.951  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      6.82e-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: 1467"
## [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
## 147 155 167 126 151 201 159 145 154 162 193 147 192 180 205
## 2011 2012
## 217 267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 77 88 83 70 99 90 91 114 121 137 96 133 113 140
## 2011 2012

```

```
## 149 177
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 67 76 75 65 84 78 78 102 108 121 87 115 100 124
## 2011 2012
## 131 158
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 53, df = 16, p-value = 8e-06
```

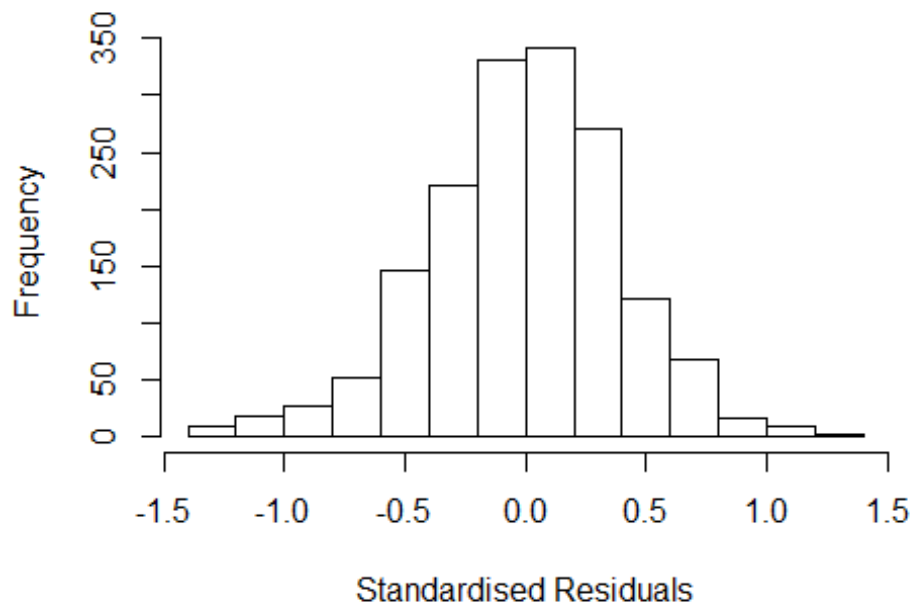


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 6.94671240523734"
## [1] "Male first author team size 2018 geometric mean: 6.12950128396187"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.30324127786858"
## [1] "Male last author team size 2018 geometric mean: 6.76438426464389"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, 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 1.067 1      1.033
## LastAuthorFemale  1.070 1      1.034
## UniqueAuthors     1.219 4      1.025
## Year              1.320 16      1.009
```

Residuals from first and last author and team size



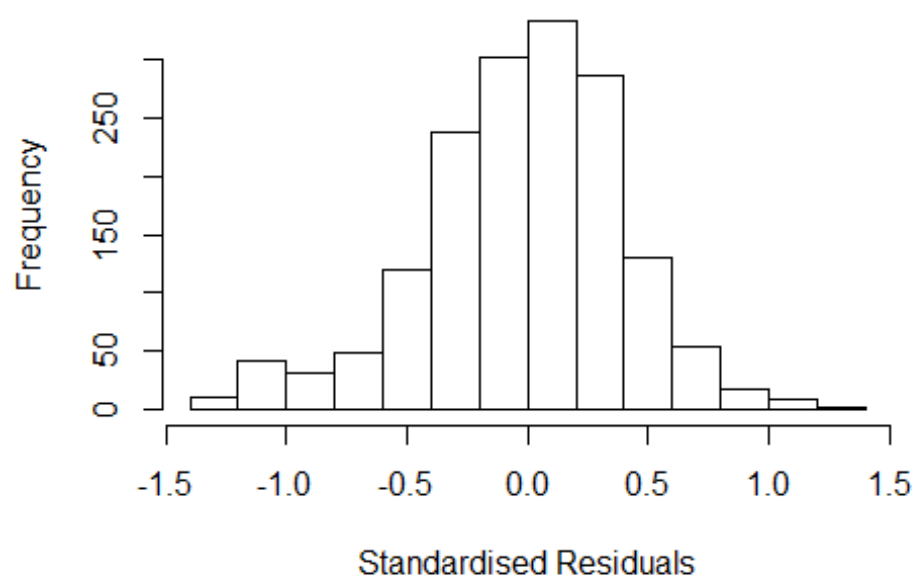
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3512 -0.2566  0.0114  0.2540  1.3172
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6362    0.0956   6.65 3.9e-11 ***
## FirstAuthorFemale1 -0.0234    0.0198  -1.18  0.238
## LastAuthorFemale1 -0.0238    0.0223  -1.06  0.287
## UniqueAuthors2    0.4916    0.0907   5.42 6.9e-08 ***
## UniqueAuthors3    0.6231    0.0858   7.26 6.0e-13 ***
## UniqueAuthors4    0.6116    0.0838   7.30 4.6e-13 ***
## UniqueAuthors5    0.7121    0.0821   8.67 < 2e-16 ***
## Year1997          0.0058    0.0805   0.07  0.943
## Year1998         -0.0458    0.0773  -0.59  0.554
## Year1999          0.0263    0.0747   0.35  0.725
```

```

## Year2000          -0.1528      0.0792   -1.93    0.054 .
## Year2001          -0.0551      0.0789   -0.70    0.485
## Year2002          -0.1141      0.0772   -1.48    0.140
## Year2003          -0.1781      0.0699   -2.55    0.011 *
## Year2004          -0.1279      0.0678   -1.89    0.059 .
## Year2005          -0.1710      0.0675   -2.53    0.011 *
## Year2006          -0.1692      0.0674   -2.51    0.012 *
## Year2007          -0.1291      0.0699   -1.85    0.065 .
## Year2008          -0.1614      0.0724   -2.23    0.026 *
## Year2009          -0.1324      0.0718   -1.84    0.065 .
## Year2010          -0.0981      0.0688   -1.43    0.154
## Year2011          -0.0620      0.0673   -0.92    0.357
## Year2012          -0.1075      0.0675   -1.59    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.149, Adjusted R-squared:  0.137
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 135 weights are ~= 1. The remaining 1491 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.140  0.867  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      6.15e-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 1.064 1      1.031
## LastAuthorFemale  1.050 1      1.025
## Year              1.113 16      1.003

```

Residuals from first and last author



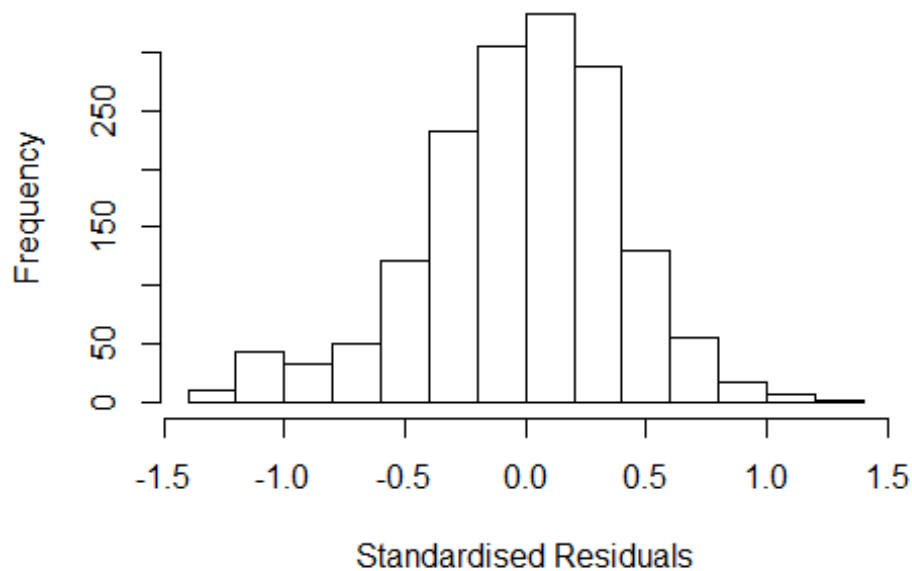
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2640 -0.2731  0.0104  0.2579  1.3598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.19496    0.06902   17.31  <2e-16 ***
## FirstAuthorFemale1 -0.00860    0.02071   -0.42    0.68
## LastAuthorFemale1 -0.03431    0.02320   -1.48    0.14
## Year1997          0.06673    0.08792    0.76    0.45
## Year1998          0.01642    0.08382    0.20    0.84
## Year1999          0.07760    0.08250    0.94    0.35
## Year2000         -0.09217    0.09994   -0.92    0.36
## Year2001          0.02643    0.08406    0.31    0.75
## Year2002         -0.02902    0.08401   -0.35    0.73
## Year2003         -0.09617    0.07635   -1.26    0.21
## Year2004         -0.03032    0.07631   -0.40    0.69
## Year2005         -0.07328    0.07559   -0.97    0.33
```

```

## Year2006      -0.07242    0.07616   -0.95    0.34
## Year2007      -0.04564    0.07807   -0.58    0.56
## Year2008      -0.08180    0.08206   -1.00    0.32
## Year2009      -0.04973    0.07786   -0.64    0.52
## Year2010       0.00593    0.07610    0.08    0.94
## Year2011       0.02571    0.07575    0.34    0.73
## Year2012      -0.00971    0.07573   -0.13    0.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0168, Adjusted R-squared:  0.00574
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1477 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.174  0.870   0.946   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      6.15e-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 1.065 1      1.032
## Year              1.065 16      1.002

```


Residuals from first author



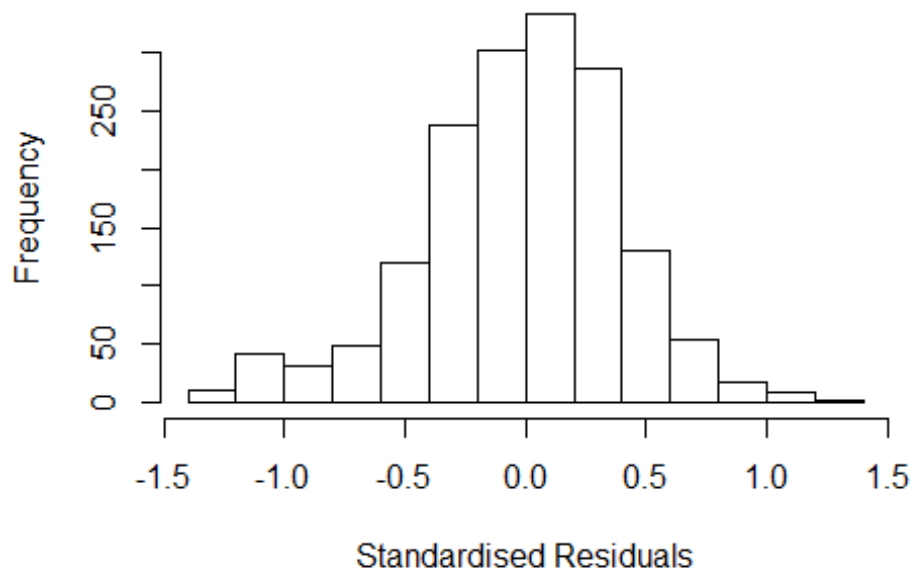
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.25397 -0.27053  0.00769  0.25954  1.36456
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.18779    0.06897   17.22  <2e-16 ***
## FirstAuthorFemale1 -0.01179    0.02077   -0.57    0.57
## Year1997          0.06618    0.08816    0.75    0.45
## Year1998          0.01593    0.08386    0.19    0.85
## Year1999          0.07610    0.08245    0.92    0.36
## Year2000         -0.09263    0.10012   -0.93    0.36
## Year2001          0.02462    0.08454    0.29    0.77
## Year2002         -0.02886    0.08416   -0.34    0.73
## Year2003         -0.09568    0.07682   -1.25    0.21
## Year2004         -0.03042    0.07664   -0.40    0.69
## Year2005         -0.07293    0.07583   -0.96    0.34
## Year2006         -0.06954    0.07631   -0.91    0.36
```

```

## Year2007          -0.04994    0.07835   -0.64    0.52
## Year2008          -0.08486    0.08219   -1.03    0.30
## Year2009          -0.04735    0.07821   -0.61    0.54
## Year2010           0.00477    0.07637    0.06    0.95
## Year2011           0.02461    0.07596    0.32    0.75
## Year2012          -0.01203    0.07600   -0.16    0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0153, Adjusted R-squared:  0.00484
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 148 weights are ~= 1. The remaining 1478 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.870  0.947  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      6.15e-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 1.051 1          1.025
## Year              1.051 16          1.002

```

Residuals from last author



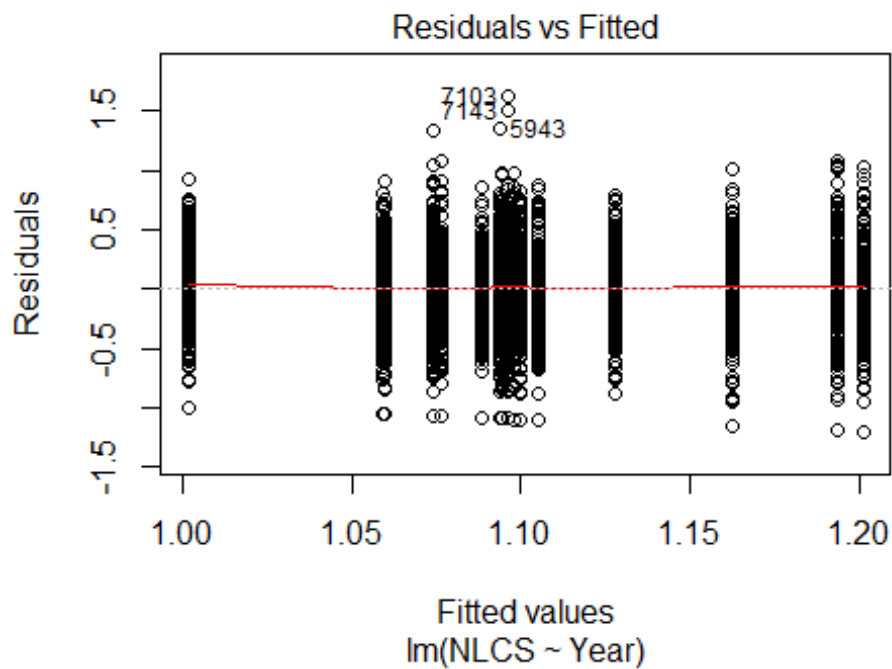
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2698 -0.2707  0.0087  0.2567  1.3643
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19124    0.06831   17.44  <2e-16 ***
## LastAuthorFemale1 -0.03529    0.02325   -1.52    0.13
## Year1997         0.06594    0.08792    0.75    0.45
## Year1998         0.01712    0.08379    0.20    0.84
## Year1999         0.07853    0.08249    0.95    0.34
## Year2000        -0.09266    0.09985   -0.93    0.35
## Year2001         0.02548    0.08411    0.30    0.76
## Year2002        -0.02957    0.08404   -0.35    0.73
## Year2003        -0.09692    0.07641   -1.27    0.20
## Year2004        -0.03151    0.07630   -0.41    0.68
## Year2005        -0.07353    0.07564   -0.97    0.33
## Year2006        -0.07283    0.07623   -0.96    0.34
```

```

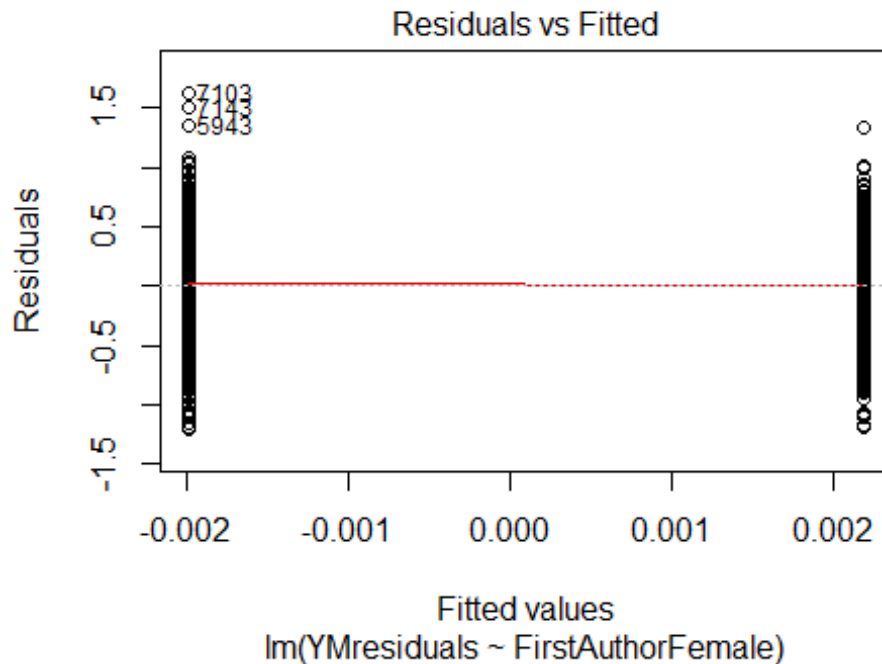
## Year2007          -0.04673      0.07815    -0.60      0.55
## Year2008          -0.08313      0.08187    -1.02      0.31
## Year2009          -0.05053      0.07784    -0.65      0.52
## Year2010           0.00524      0.07612     0.07      0.95
## Year2011           0.02508      0.07578     0.33      0.74
## Year2012          -0.01061      0.07580    -0.14      0.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00626
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1468 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.870  0.945  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      6.15e-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: 1626"
## [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
##  407  392  389  352  430  393  398  290  324  321  338  349  332  401  372
## 2011 2012
##  372  377
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  240  209  227  233  203  173  261  185  224  234  242  229  229  281  265
## 2011 2012

```

```
## 278 258
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 217 186 208 212 178 158 229 163 192 204 214 196 204 247 232
## 2011 2012
## 241 231
## [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
```

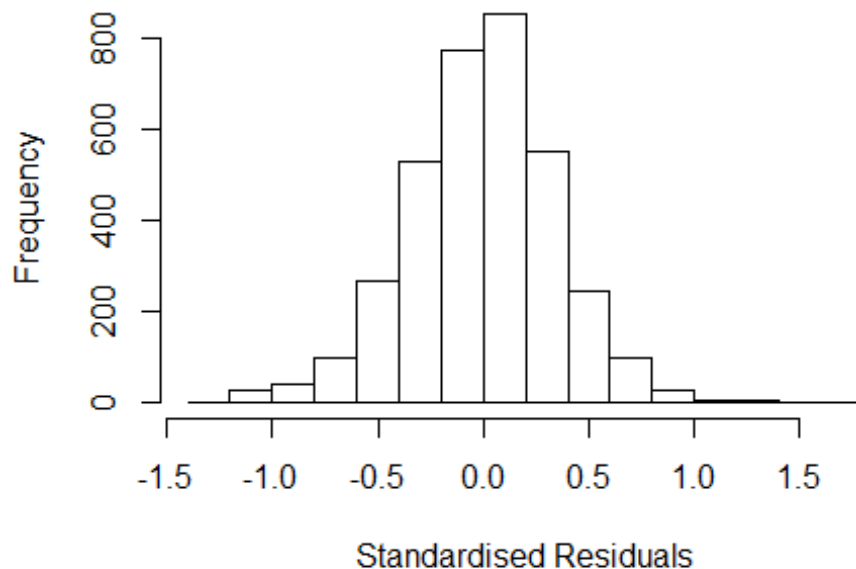


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 30, df = 1, p-value = 5e-08
```



```
## [1] "Female first author team size 2018 geometric mean: 5.81888526707223"
## [1] "Male first author team size 2018 geometric mean: 5.40952675778963"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4600, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.13412377269765"
## [1] "Male last author team size 2018 geometric mean: 5.7984313762336"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3000, 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 1.034 1      1.017
## LastAuthorFemale  1.043 1      1.021
## UniqueAuthors    1.176 4      1.020
## Year              1.216 16      1.006
```

Residuals from first and last author and team size



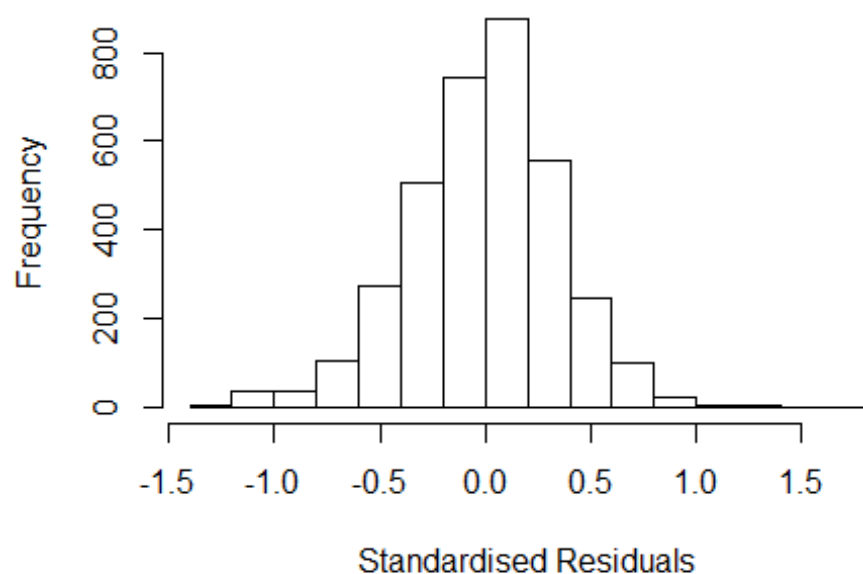
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32036 -0.22407 0.00666 0.21541 1.74907
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07033 0.04703 22.76 < 2e-16 ***
## FirstAuthorFemale1 -0.01676 0.01170 -1.43 0.15211
## LastAuthorFemale1 -0.04474 0.01398 -3.20 0.00138 **
## UniqueAuthors2 0.10185 0.04197 2.43 0.01528 *
## UniqueAuthors3 0.15579 0.04041 3.85 0.00012 ***
## UniqueAuthors4 0.18781 0.04051 4.64 3.7e-06 ***
## UniqueAuthors5 0.26679 0.03936 6.78 1.4e-11 ***
## Year1997 -0.00997 0.04023 -0.25 0.80426
## Year1998 -0.04612 0.04018 -1.15 0.25114
## Year1999 -0.11482 0.03788 -3.03 0.00246 **
```

```

## Year2000      -0.19304      0.04236      -4.56      5.4e-06 ***
## Year2001      -0.17970      0.03973      -4.52      6.3e-06 ***
## Year2002      -0.15460      0.03681      -4.20      2.7e-05 ***
## Year2003      -0.17548      0.03824      -4.59      4.6e-06 ***
## Year2004      -0.18540      0.03693      -5.02      5.4e-07 ***
## Year2005      -0.16578      0.03802      -4.36      1.3e-05 ***
## Year2006      -0.16962      0.03739      -4.54      5.9e-06 ***
## Year2007      -0.18615      0.03874      -4.81      1.6e-06 ***
## Year2008      -0.16108      0.03809      -4.23      2.4e-05 ***
## Year2009      -0.12709      0.03585      -3.55      0.00040 ***
## Year2010      -0.16051      0.03694      -4.35      1.4e-05 ***
## Year2011      -0.15494      0.03834      -4.04      5.4e-05 ***
## Year2012      -0.15415      0.03908      -3.94      8.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.327
## Multiple R-squared:  0.0642, Adjusted R-squared:  0.0583
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(2977,2992) are outliers with |weight| = 0 ( < 2.8e-05);
## 298 weights are ~= 1. The remaining 3212 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0415 0.8690 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      2.85e-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 1.028 1      1.014
## LastAuthorFemale 1.039 1      1.019
## Year      1.049 16      1.002

```


Residuals from first and last author



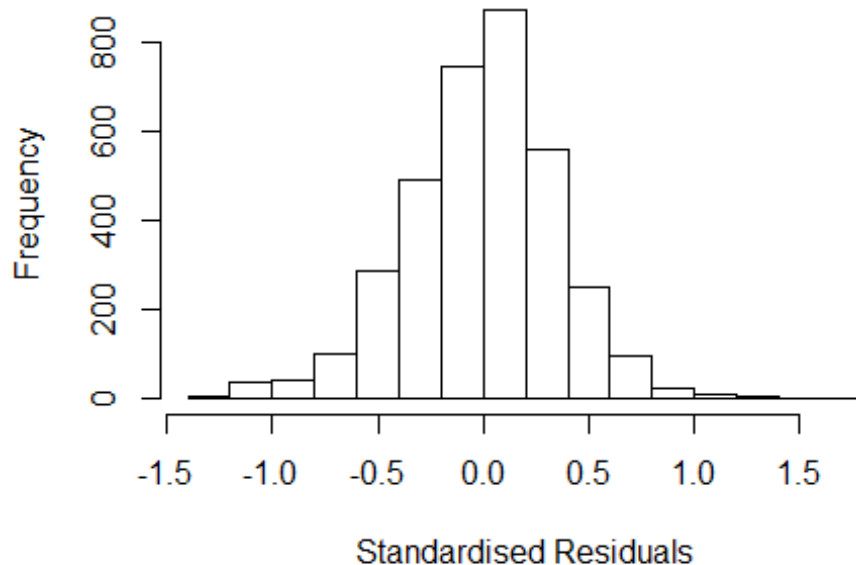
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2299 -0.2287 0.0132 0.2136 1.6460
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22754 0.03022 40.62 < 2e-16 ***
## FirstAuthorFemale1 -0.00618 0.01192 -0.52 0.60387
## LastAuthorFemale1 -0.05183 0.01429 -3.63 0.00029 ***
## Year1997 0.00233 0.04090 0.06 0.95464
## Year1998 -0.03638 0.04004 -0.91 0.36364
## Year1999 -0.09403 0.03788 -2.48 0.01311 *
## Year2000 -0.16665 0.04207 -3.96 7.6e-05 ***
## Year2001 -0.14291 0.04032 -3.54 0.00040 ***
## Year2002 -0.11006 0.03682 -2.99 0.00282 **
## Year2003 -0.14389 0.03908 -3.68 0.00024 ***
## Year2004 -0.13855 0.03654 -3.79 0.00015 ***
## Year2005 -0.12456 0.03766 -3.31 0.00095 ***
```

```

## Year2006          -0.12154    0.03687   -3.30  0.00099 ***
## Year2007          -0.13448    0.03883   -3.46  0.00054 ***
## Year2008          -0.10719    0.03765   -2.85  0.00444 **
## Year2009          -0.07578    0.03532   -2.15  0.03195 *
## Year2010          -0.10572    0.03672   -2.88  0.00401 **
## Year2011          -0.10281    0.03833   -2.68  0.00735 **
## Year2012          -0.10028    0.03845   -2.61  0.00915 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.0215, Adjusted R-squared:  0.0164
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2977 is an outlier with |weight| = 0 ( < 2.8e-05);
## 273 weights are ~= 1. The remaining 3238 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8680 0.9520 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      2.85e-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 1.02 1          1.010
## Year              1.02 16          1.001

```

Residuals from first author



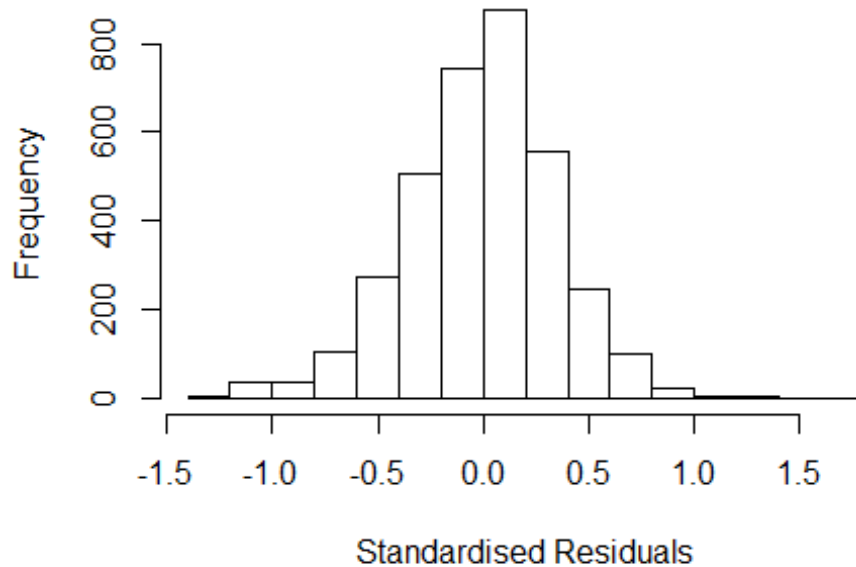
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2209 -0.2250 0.0111 0.2158 1.6053
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21594 0.02983 40.76 < 2e-16 ***
## FirstAuthorFemale1 -0.00893 0.01191 -0.75 0.45319
## Year1997 0.00492 0.04098 0.12 0.90441
## Year1998 -0.03329 0.03989 -0.83 0.40410
## Year1999 -0.09293 0.03785 -2.46 0.01412 *
## Year2000 -0.16288 0.04205 -3.87 0.00011 ***
## Year2001 -0.13873 0.04040 -3.43 0.00060 ***
## Year2002 -0.10698 0.03679 -2.91 0.00366 **
## Year2003 -0.14288 0.03902 -3.66 0.00025 ***
## Year2004 -0.13681 0.03641 -3.76 0.00017 ***
## Year2005 -0.12045 0.03766 -3.20 0.00140 **
## Year2006 -0.11932 0.03678 -3.24 0.00119 **
```

```

## Year2007          -0.13403      0.03868      -3.47  0.00054 ***
## Year2008          -0.10632      0.03760      -2.83  0.00472 **
## Year2009          -0.07584      0.03519      -2.16  0.03123 *
## Year2010          -0.10523      0.03668      -2.87  0.00415 **
## Year2011          -0.10519      0.03828      -2.75  0.00603 **
## Year2012          -0.10048      0.03826      -2.63  0.00868 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.333
## Multiple R-squared:  0.0175, Adjusted R-squared:  0.0127
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2977 is an outlier with |weight| = 0 ( < 2.8e-05);
## 281 weights are ~= 1. The remaining 3230 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8660 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          2.85e-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 1.03 1          1.015
## Year          1.03 16          1.001

```

Residuals from last author



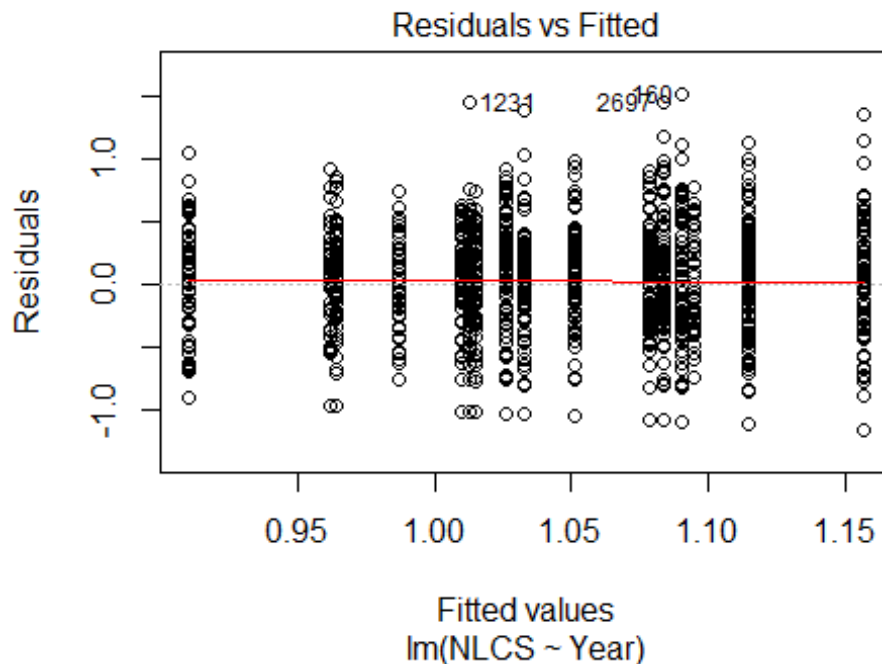
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2275 -0.2292 0.0139 0.2138 1.6494
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2253 0.0299 41.02 < 2e-16 ***
## LastAuthorFemale1 -0.0523 0.0142 -3.67 0.00024 ***
## Year1997 0.0022 0.0409 0.05 0.95708
## Year1998 -0.0370 0.0400 -0.92 0.35518
## Year1999 -0.0944 0.0379 -2.49 0.01277 *
## Year2000 -0.1672 0.0420 -3.98 7.1e-05 ***
## Year2001 -0.1436 0.0403 -3.56 0.00037 ***
## Year2002 -0.1106 0.0368 -3.01 0.00267 **
## Year2003 -0.1450 0.0390 -3.72 0.00021 ***
## Year2004 -0.1396 0.0365 -3.83 0.00013 ***
## Year2005 -0.1252 0.0376 -3.33 0.00089 ***
## Year2006 -0.1223 0.0369 -3.32 0.00092 ***
```

```

## Year2007          -0.1356      0.0388   -3.50  0.00048 ***
## Year2008          -0.1078      0.0376   -2.86  0.00421 **
## Year2009          -0.0768      0.0352   -2.18  0.02943 *
## Year2010          -0.1064      0.0367   -2.90  0.00377 **
## Year2011          -0.1038      0.0382   -2.71  0.00667 **
## Year2012          -0.1015      0.0384   -2.64  0.00824 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.0214, Adjusted R-squared:  0.0166
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2977 is an outlier with |weight| = 0 ( < 2.8e-05);
## 274 weights are ~= 1. The remaining 3237 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0004 0.8680 0.9530 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          2.85e-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: 3512"
## [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
## 164 121 116 99 190 137 118 90 106 114 117 116 112 137 149
## 2011 2012
## 150 141
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 46 60 58 63 47 75 45 75 90 86 84 75 92 94

```

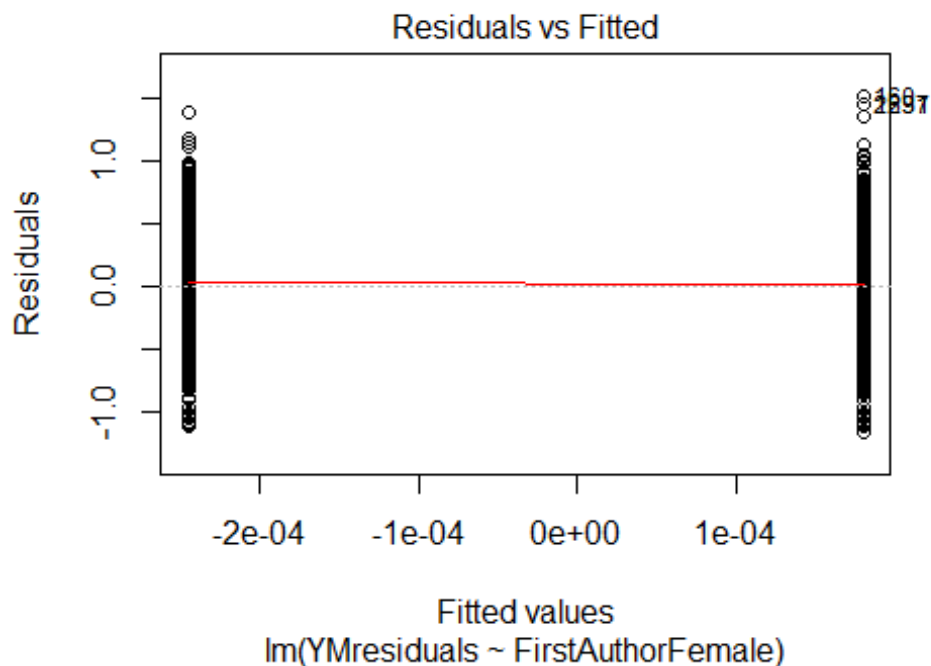
```
## 2011 2012
## 115 103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 42 54 49 55 42 67 39 69 80 76 77 66 82 83
## 2011 2012
## 99 96
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.05
```



```
##
## 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: 4.86834709482937"
## [1] "Male first author team size 2018 geometric mean: 5.19837917165599"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

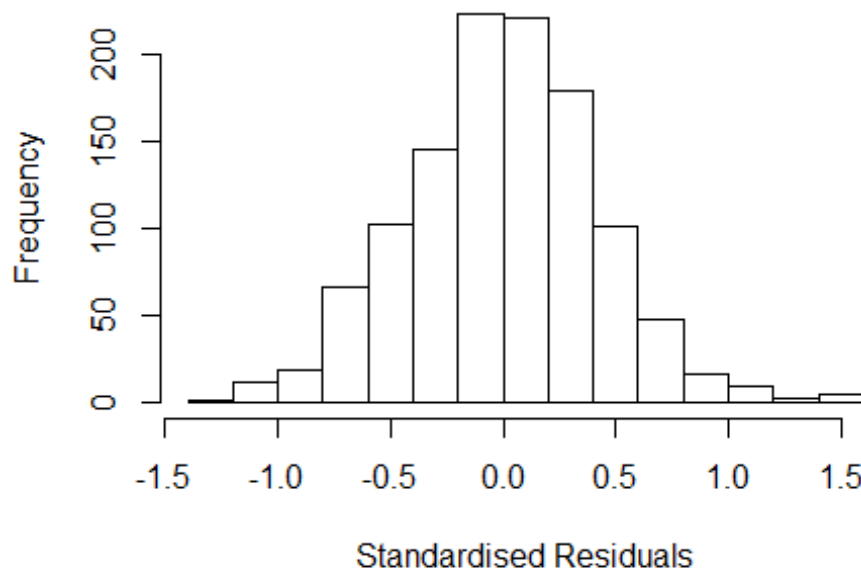
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.35703943921425"
## [1] "Male last author team size 2018 geometric mean: 4.95575197119143"

## 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 = 450, 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  1.080 1      1.039
## UniqueAuthors    1.453 4      1.048
## Year              1.558 16     1.014
```


Residuals from first and last author and team size



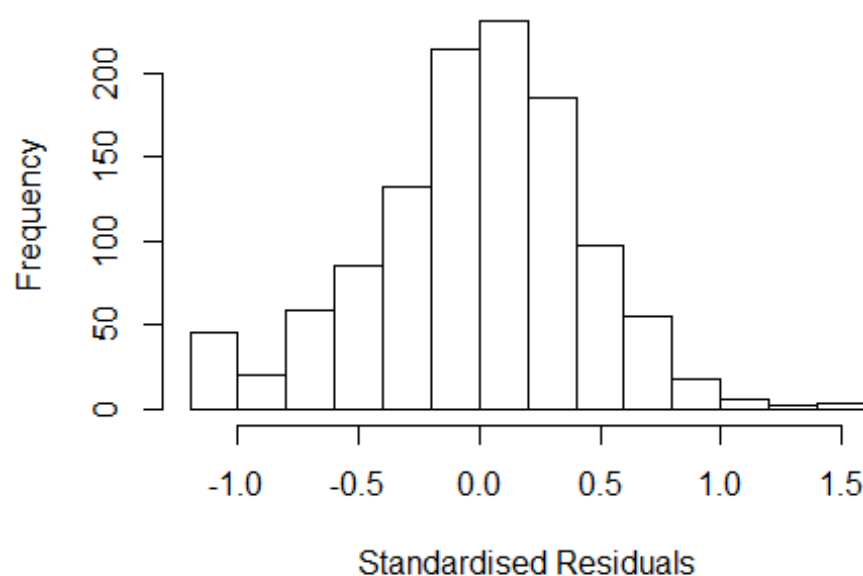
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22011 -0.26377  0.00161  0.26955  1.50807
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.71679    0.09989   7.18 1.3e-12 ***
## FirstAuthorFemale1 -0.01883    0.02569  -0.73 0.46371
## LastAuthorFemale1 -0.07151    0.03342  -2.14 0.03260 *
## UniqueAuthors2     0.30088    0.08565   3.51 0.00046 ***
## UniqueAuthors3     0.38153    0.08260   4.62 4.3e-06 ***
## UniqueAuthors4     0.37613    0.08351   4.50 7.4e-06 ***
## UniqueAuthors5     0.52216    0.08074   6.47 1.5e-10 ***
## Year1997         -0.01212    0.09036  -0.13 0.89335
## Year1998          0.01214    0.09054   0.13 0.89333
## Year1999          0.00168    0.08853   0.02 0.98491
```

```

## Year2000      -0.11309    0.08910   -1.27  0.20462
## Year2001      -0.08864    0.08948   -0.99  0.32211
## Year2002      -0.04965    0.07643   -0.65  0.51605
## Year2003       0.02180    0.08566    0.25  0.79920
## Year2004      -0.02802    0.07795   -0.36  0.71928
## Year2005      -0.08414    0.07762   -1.08  0.27856
## Year2006      -0.10447    0.07824   -1.34  0.18207
## Year2007       0.03134    0.08697    0.36  0.71867
## Year2008       0.00155    0.08775    0.02  0.98594
## Year2009      -0.06182    0.08019   -0.77  0.44090
## Year2010      -0.06117    0.07804   -0.78  0.43334
## Year2011      -0.02810    0.07979   -0.35  0.72477
## Year2012      -0.04586    0.07977   -0.57  0.56543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.126, Adjusted R-squared:  0.109
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1041 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.117  0.848  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           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 1.051 1           1.025
## LastAuthorFemale  1.111 1           1.054
## Year              1.167 16           1.005

```

Residuals from first and last author



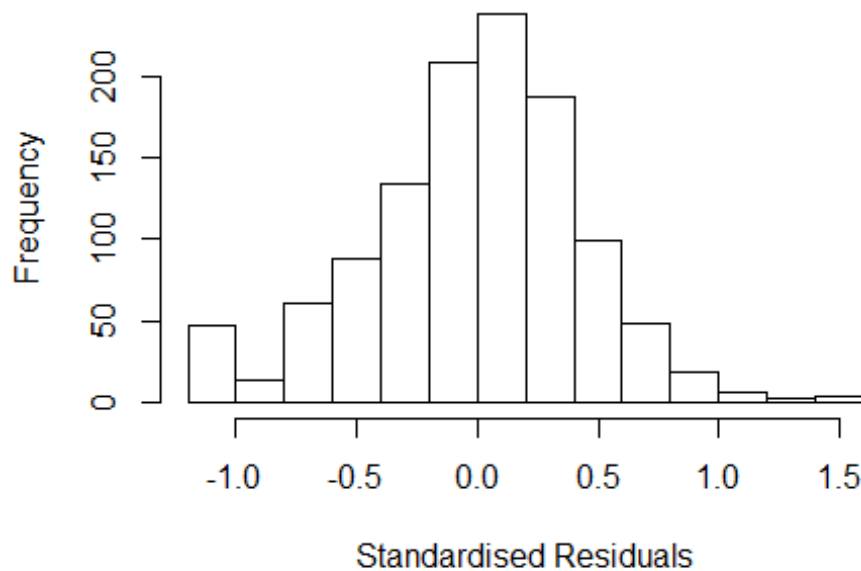
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1771 -0.2865 0.0258 0.2589 1.5272
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07581 0.06545 16.44 <2e-16 ***
## FirstAuthorFemale1 -0.00369 0.02640 -0.14 0.889
## LastAuthorFemale1 -0.06520 0.03765 -1.73 0.084 .
## Year1997 -0.04324 0.09128 -0.47 0.636
## Year1998 0.05774 0.09073 0.64 0.525
## Year1999 -0.00686 0.08820 -0.08 0.938
## Year2000 -0.09198 0.09298 -0.99 0.323
## Year2001 -0.07624 0.08664 -0.88 0.379
## Year2002 -0.00862 0.07874 -0.11 0.913
## Year2003 0.08193 0.09181 0.89 0.372
## Year2004 -0.02435 0.07994 -0.30 0.761
## Year2005 -0.02935 0.07902 -0.37 0.710
```

```

## Year2006      -0.08397    0.07932   -1.06    0.290
## Year2007      0.10128    0.08762    1.16    0.248
## Year2008      0.07096    0.08886    0.80    0.425
## Year2009      0.01065    0.08134    0.13    0.896
## Year2010      0.01961    0.07777    0.25    0.801
## Year2011     -0.00181    0.08267   -0.02    0.983
## Year2012      0.01290    0.08077    0.16    0.873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.00258
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 1071 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.125  0.858  0.951  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      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 1.048 1      1.024
## Year              1.048 16      1.001

```

Residuals from first author



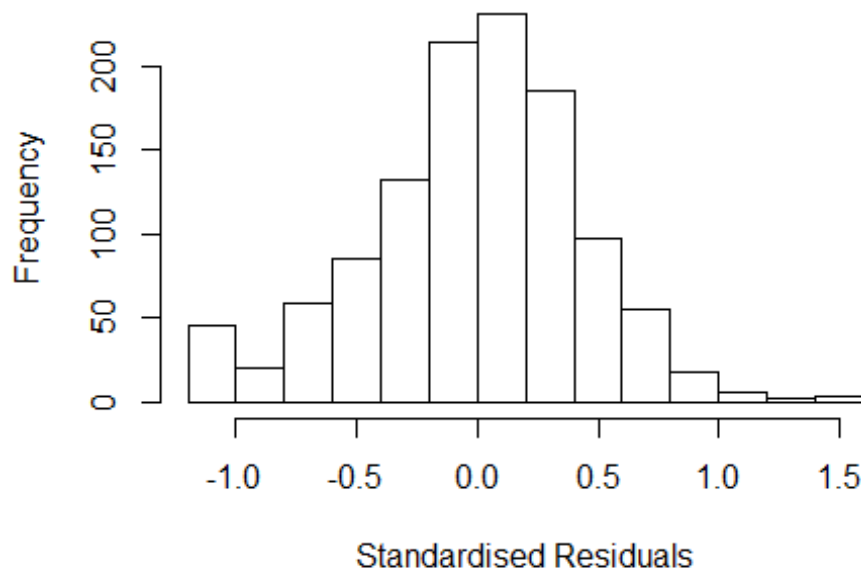
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1638 -0.2772 0.0218 0.2654 1.5373
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06569 0.06523 16.34 <2e-16 ***
## FirstAuthorFemale1 -0.00734 0.02647 -0.28 0.78
## Year1997 -0.05816 0.08900 -0.65 0.51
## Year1998 0.05873 0.09034 0.65 0.52
## Year1999 -0.01066 0.08763 -0.12 0.90
## Year2000 -0.09430 0.09285 -1.02 0.31
## Year2001 -0.07287 0.08638 -0.84 0.40
## Year2002 -0.01154 0.07823 -0.15 0.88
## Year2003 0.07926 0.09134 0.87 0.39
## Year2004 -0.02315 0.07919 -0.29 0.77
## Year2005 -0.02828 0.07881 -0.36 0.72
## Year2006 -0.08240 0.07893 -1.04 0.30
```

```

## Year2007          0.09813    0.08639    1.14    0.26
## Year2008          0.07550    0.08824    0.86    0.39
## Year2009          0.00638    0.08045    0.08    0.94
## Year2010          0.02147    0.07716    0.28    0.78
## Year2011         -0.00393    0.08178   -0.05    0.96
## Year2012          0.01277    0.07966    0.16    0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0145, Adjusted R-squared:  -0.000279
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 1061 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.119  0.854  0.953   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      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 1.109 1      1.053
## Year              1.109 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
## -1.1755 -0.2869 0.0274 0.2579 1.5287
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07434 0.06376 16.85 <2e-16 ***
## LastAuthorFemale1 -0.06547 0.03774 -1.73 0.083 .
## Year1997 -0.04272 0.09115 -0.47 0.639
## Year1998 0.05769 0.09074 0.64 0.525
## Year1999 -0.00673 0.08820 -0.08 0.939
## Year2000 -0.09197 0.09306 -0.99 0.323
## Year2001 -0.07686 0.08680 -0.89 0.376
## Year2002 -0.00847 0.07877 -0.11 0.914
## Year2003 0.08143 0.09196 0.89 0.376
## Year2004 -0.02471 0.07999 -0.31 0.757
## Year2005 -0.02969 0.07901 -0.38 0.707
## Year2006 -0.08405 0.07942 -1.06 0.290
```

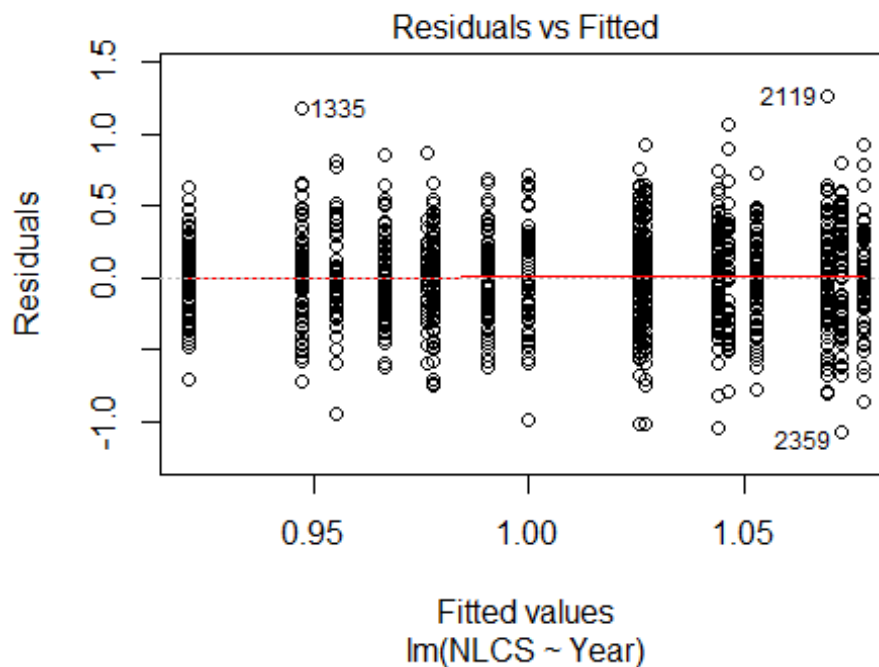
```

## Year2007      0.10114      0.08781      1.15      0.250
## Year2008      0.07119      0.08883      0.80      0.423
## Year2009      0.01063      0.08140      0.13      0.896
## Year2010      0.01951      0.07786      0.25      0.802
## Year2011     -0.00169      0.08268     -0.02      0.984
## Year2012      0.01298      0.08082      0.16      0.872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.00345
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 1069 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.124  0.857  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      8.68e-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 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
## 109 110 114 98 93 128 137 116 114 111 123 124 123 138 127
## 2011 2012
## 115 132
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 57 60 46 37 55 83 84 78 74 90 89 87 85 82
## 2011 2012

```



```
## 71 87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 50 49 55 40 34 51 75 69 67 63 84 81 80 75 79
## 2011 2012
## 66 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 = 33, df = 16, p-value = 0.007
```



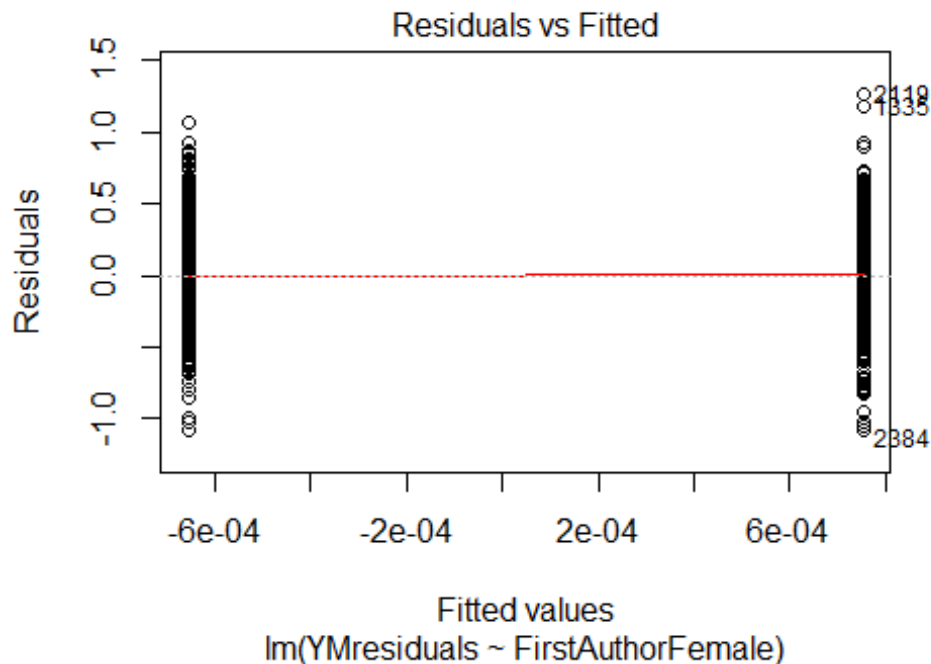
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 4.71703250320574"
## [1] "Male first author team size 2018 geometric mean: 4.52447041526636"

## 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 = 590, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.48049660376863"
## [1] "Male last author team size 2018 geometric mean: 4.76133303188226"

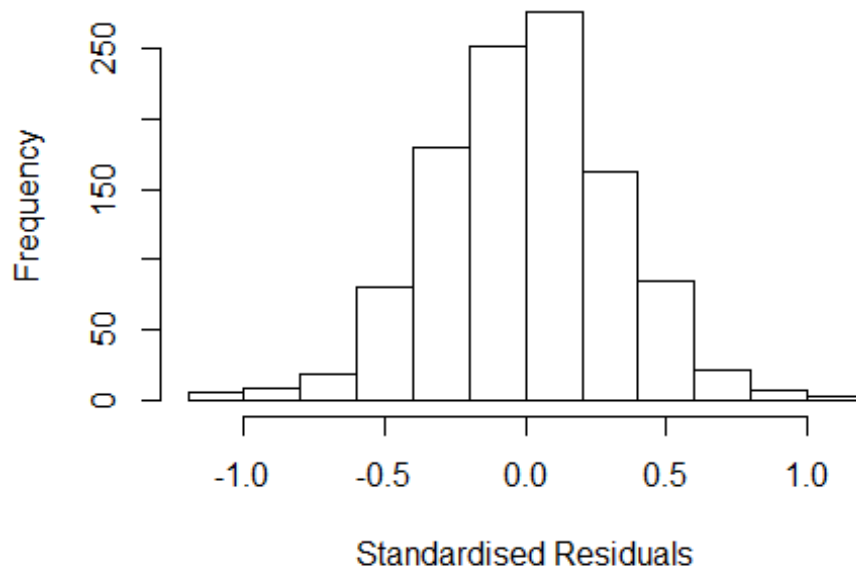
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 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	1.129	1	1.063
LastAuthorFemale	1.164	1	1.079
UniqueAuthors	1.515	4	1.053
Year	1.614	16	1.015

Residuals from first and last author and team size



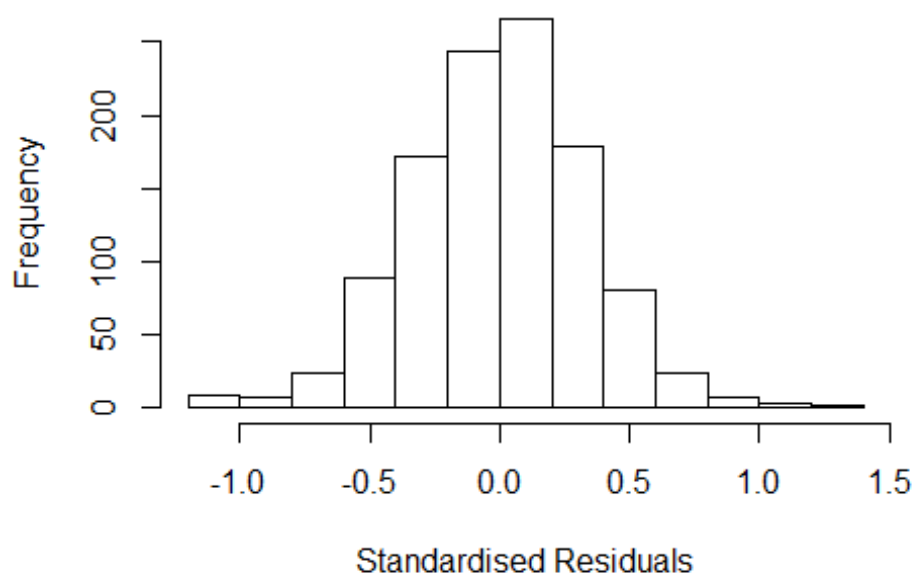
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05243 -0.21013  0.00473  0.20614  1.17728
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.981193   0.060844   16.13 < 2e-16 ***
## FirstAuthorFemale1 -0.021793   0.020864   -1.04  0.29647
## LastAuthorFemale1 -0.041770   0.022528   -1.85  0.06399 .
## UniqueAuthors2     0.043236   0.041999    1.03  0.30349
## UniqueAuthors3     0.058037   0.040001    1.45  0.14710
## UniqueAuthors4     0.040028   0.042356    0.95  0.34484
## UniqueAuthors5     0.136676   0.039603    3.45  0.00058 ***
## Year1997           0.070466   0.075392    0.93  0.35018
## Year1998           0.000235   0.073460    0.00  0.99744
## Year1999          -0.113731   0.075600   -1.50  0.13278
```

```

## Year2000      -0.040813    0.073058   -0.56   0.57652
## Year2001      0.033557    0.070223    0.48   0.63284
## Year2002     -0.054224    0.065847   -0.82   0.41041
## Year2003     -0.081762    0.065684   -1.24   0.21348
## Year2004     -0.040166    0.067903   -0.59   0.55430
## Year2005     -0.052459    0.068652   -0.76   0.44496
## Year2006      0.027998    0.068001    0.41   0.68063
## Year2007      0.019454    0.069632    0.28   0.78000
## Year2008     -0.020871    0.072144   -0.29   0.77241
## Year2009     -0.025240    0.069893   -0.36   0.71808
## Year2010      0.035850    0.068446    0.52   0.60054
## Year2011      0.044301    0.074196    0.60   0.55058
## Year2012      0.007403    0.072972    0.10   0.91921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.317
## Multiple R-squared:  0.0454, Adjusted R-squared:  0.0259
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 1013 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  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      9.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.093 1      1.045
## LastAuthorFemale  1.104 1      1.051
## Year              1.122 16      1.004

```

Residuals from first and last author



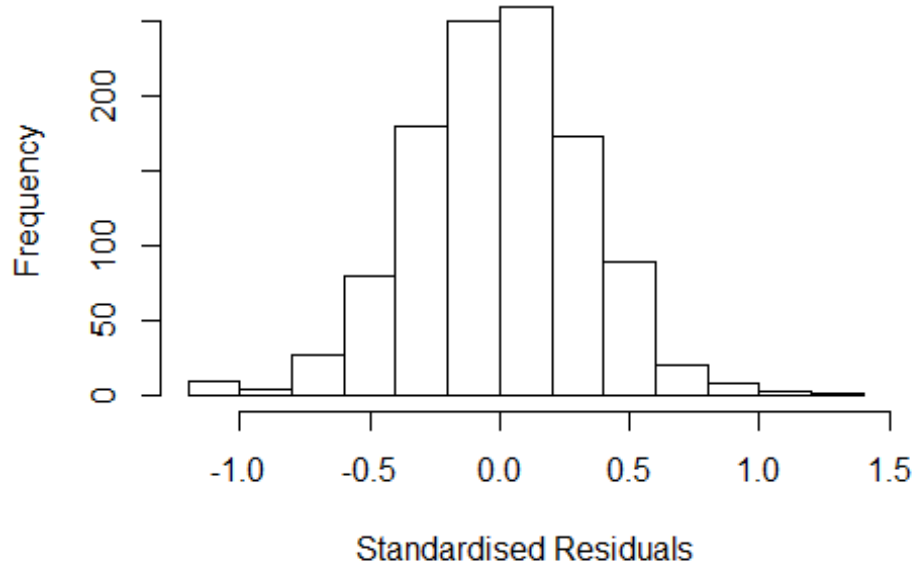
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1050 -0.2233  0.0114  0.2155  1.2366
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02925    0.05571   18.48  <2e-16 ***
## FirstAuthorFemale1 -0.01977    0.02091   -0.95   0.345
## LastAuthorFemale1 -0.04997    0.02231   -2.24   0.025 *
## Year1997         0.08336    0.07545    1.10   0.269
## Year1998         0.01100    0.07228    0.15   0.879
## Year1999        -0.09299    0.07324   -1.27   0.204
## Year2000        -0.02660    0.07012   -0.38   0.705
## Year2001         0.03714    0.06930    0.54   0.592
## Year2002        -0.03835    0.06417   -0.60   0.550
## Year2003        -0.07139    0.06516   -1.10   0.273
## Year2004        -0.01020    0.06608   -0.15   0.877
## Year2005        -0.04216    0.06689   -0.63   0.529
```

```

## Year2006          0.06291    0.06559    0.96    0.338
## Year2007          0.04719    0.06785    0.70    0.487
## Year2008          0.01216    0.06884    0.18    0.860
## Year2009         -0.00149    0.06852   -0.02    0.983
## Year2010          0.06515    0.06637    0.98    0.327
## Year2011          0.07579    0.07413    1.02    0.307
## Year2012          0.05500    0.07005    0.79    0.433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.0278, Adjusted R-squared:  0.0117
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 998 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.869  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      9.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.056 1      1.028
## Year              1.056 16      1.002

```

Residuals from first author



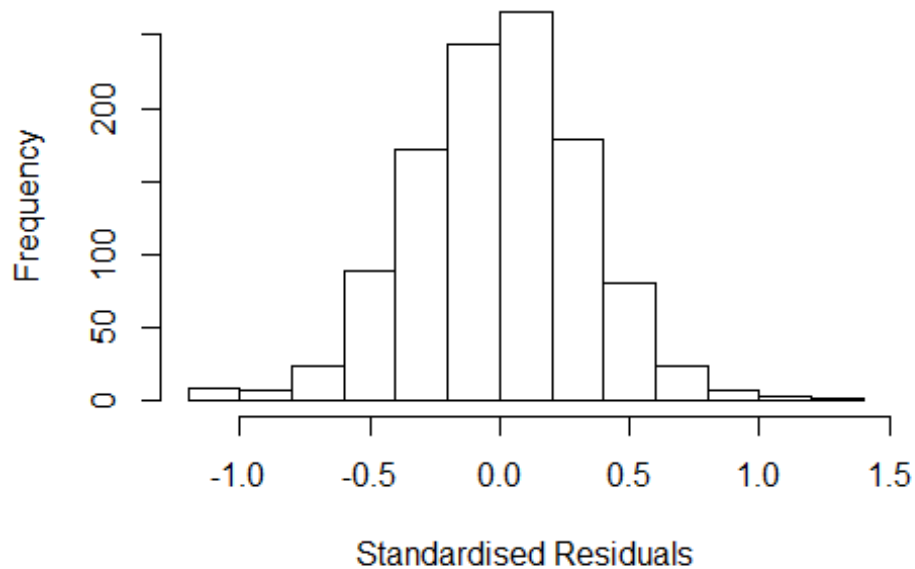
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.09485 -0.21958 0.00392 0.21426 1.24904
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02504 0.05546 18.48 <2e-16 ***
## FirstAuthorFemale1 -0.02666 0.02067 -1.29 0.20
## Year1997 0.07128 0.07470 0.95 0.34
## Year1998 0.00673 0.07188 0.09 0.93
## Year1999 -0.10010 0.07290 -1.37 0.17
## Year2000 -0.03164 0.06916 -0.46 0.65
## Year2001 0.02907 0.06967 0.42 0.68
## Year2002 -0.04113 0.06397 -0.64 0.52
## Year2003 -0.08200 0.06490 -1.26 0.21
## Year2004 -0.01586 0.06603 -0.24 0.81
## Year2005 -0.04419 0.06721 -0.66 0.51
## Year2006 0.05311 0.06574 0.81 0.42
```

```

## Year2007          0.04060      0.06746      0.60      0.55
## Year2008          0.00471      0.06880      0.07      0.95
## Year2009         -0.00554      0.06885     -0.08      0.94
## Year2010          0.05693      0.06655      0.86      0.39
## Year2011          0.06982      0.07446      0.94      0.35
## Year2012          0.04731      0.06970      0.68      0.50
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.322
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.00756
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 1015 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0984 0.8700 0.9510 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      9.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.069 1      1.034
## Year              1.069 16      1.002

```


Residuals from last author



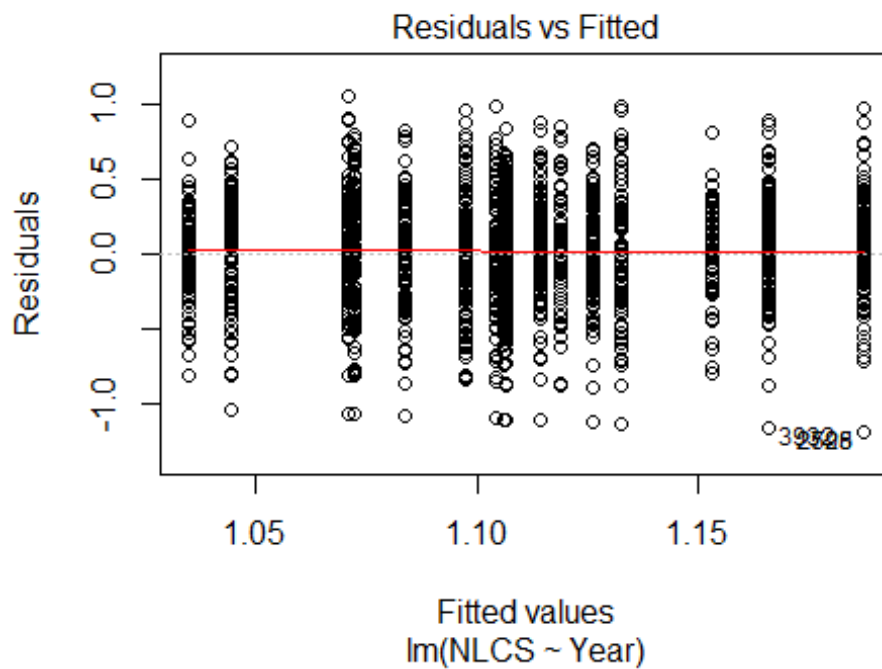
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.09305 -0.22807 0.00954 0.21298 1.24770
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02323 0.05591 18.30 <2e-16 ***
## LastAuthorFemale1 -0.05318 0.02201 -2.42 0.016 *
## Year1997 0.08251 0.07597 1.09 0.278
## Year1998 0.00766 0.07289 0.11 0.916
## Year1999 -0.09409 0.07404 -1.27 0.204
## Year2000 -0.03039 0.07083 -0.43 0.668
## Year2001 0.03544 0.06959 0.51 0.611
## Year2002 -0.04110 0.06477 -0.63 0.526
## Year2003 -0.07611 0.06546 -1.16 0.245
## Year2004 -0.01299 0.06648 -0.20 0.845
## Year2005 -0.04702 0.06706 -0.70 0.483
## Year2006 0.05655 0.06563 0.86 0.389
```

```

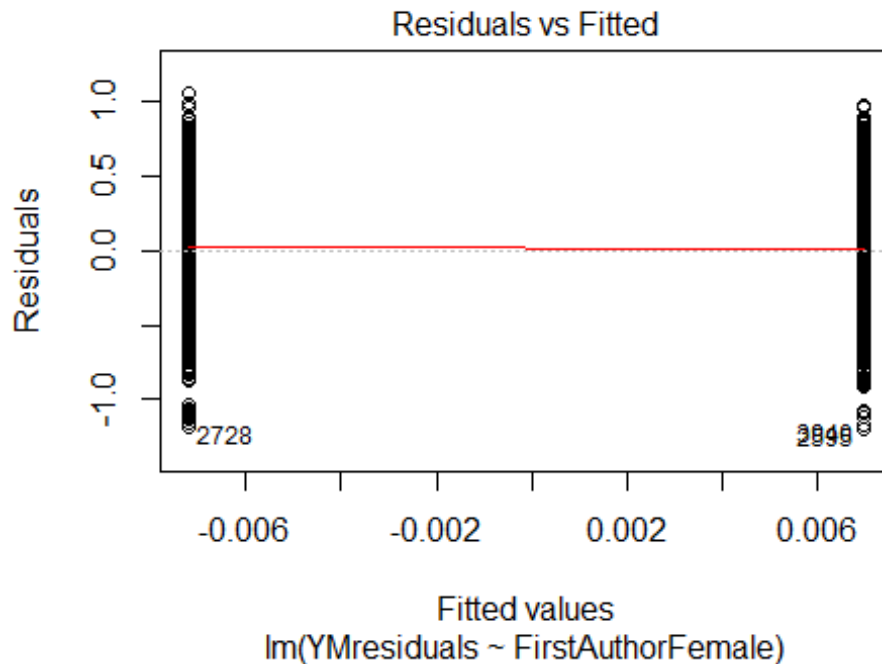
## Year2007      0.04288      0.06823      0.63      0.530
## Year2008      0.00780      0.06916      0.11      0.910
## Year2009     -0.00560      0.06878     -0.08      0.935
## Year2010      0.06007      0.06651      0.90      0.367
## Year2011      0.06982      0.07421      0.94      0.347
## Year2012      0.04982      0.07046      0.71      0.480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.323
## Multiple R-squared:  0.0268, Adjusted R-squared:  0.0115
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1003 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.872  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      9.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1100"
## [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
## 210 213 171 171 175 219 215 150 174 177 191 202 190 198 210
## 2011 2012
## 188 215
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 99 98 75 74 55 72 118 93 107 111 121 120 116 114 114
## 2011 2012

```

```
## 124 129
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 91 70 68 48 64 104 71 87 93 106 107 101 95 100
## 2011 2012
## 107 113
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```

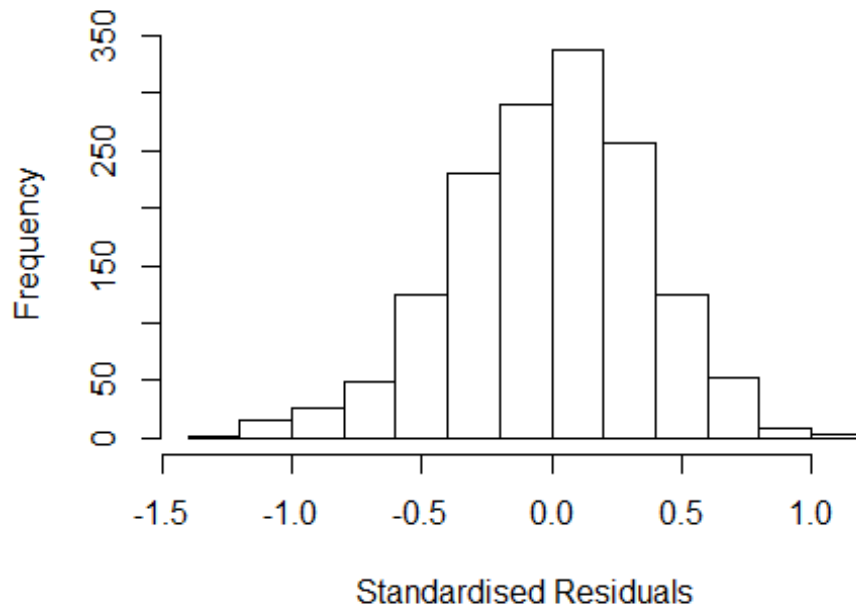


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.1, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 4.9217202221825"
## [1] "Male first author team size 2018 geometric mean: 5.51464967736117"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.96161530176954"
## [1] "Male last author team size 2018 geometric mean: 5.27543271376378"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, 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 1.099 1      1.048
## LastAuthorFemale  1.049 1      1.024
## UniqueAuthors    1.363 4      1.039
## Year              1.487 16     1.012
```

Residuals from first and last author and team size



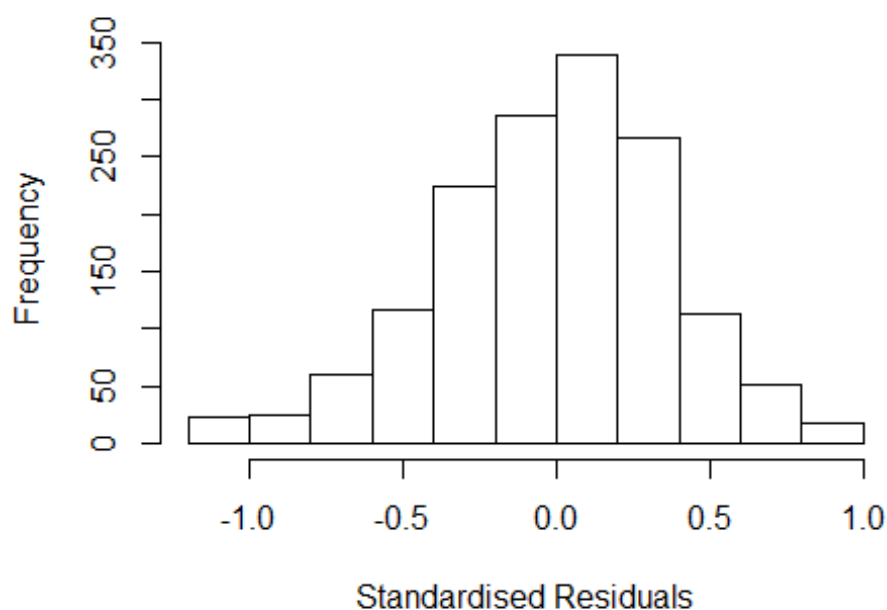
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2437 -0.2481 0.0129 0.2385 1.0454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00059 0.06121 16.35 < 2e-16 ***
## FirstAuthorFemale1 0.00948 0.01985 0.48 0.6330
## LastAuthorFemale1 -0.01004 0.02283 -0.44 0.6603
## UniqueAuthors2 0.06263 0.05339 1.17 0.2410
## UniqueAuthors3 0.11590 0.05206 2.23 0.0262 *
## UniqueAuthors4 0.13423 0.05199 2.58 0.0099 **
## UniqueAuthors5 0.24315 0.04947 4.92 9.8e-07 ***
## Year1997 0.01598 0.06405 0.25 0.8030
## Year1998 -0.01441 0.06324 -0.23 0.8198
## Year1999 -0.02949 0.06351 -0.46 0.6425
```

```

## Year2000      0.01698    0.06759    0.25    0.8017
## Year2001     -0.07661    0.06021   -1.27    0.2034
## Year2002     -0.02702    0.05725   -0.47    0.6370
## Year2003     -0.02212    0.05729   -0.39    0.6995
## Year2004     -0.09013    0.05843   -1.54    0.1232
## Year2005     -0.02893    0.05550   -0.52    0.6022
## Year2006     -0.03121    0.05412   -0.58    0.5642
## Year2007      0.02300    0.05526    0.42    0.6773
## Year2008     -0.14457    0.05694   -2.54    0.0112 *
## Year2009     -0.04586    0.05959   -0.77    0.4416
## Year2010     -0.08743    0.06049   -1.45    0.1486
## Year2011     -0.06275    0.06113   -1.03    0.3048
## Year2012     -0.00817    0.05615   -0.15    0.8844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0549, Adjusted R-squared:  0.041
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1395 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.211  0.876  0.950  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      6.58e-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 1.099 1      1.048
## LastAuthorFemale  1.035 1      1.017
## Year              1.133 16      1.004

```

Residuals from first and last author



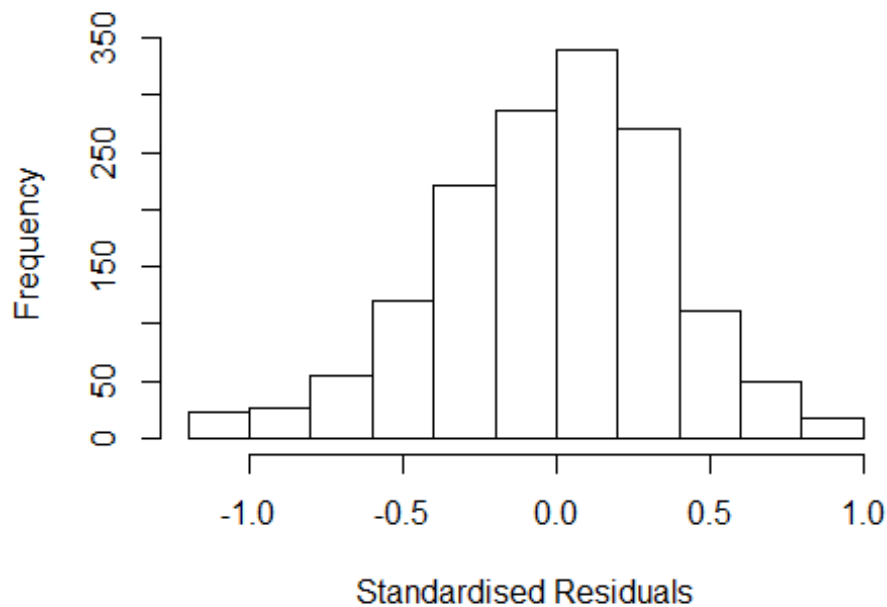
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.192 -0.254 0.014 0.247 0.976
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11706 0.04554 24.53 <2e-16 ***
## FirstAuthorFemale1 0.01645 0.02033 0.81 0.42
## LastAuthorFemale1 -0.01873 0.02299 -0.81 0.42
## Year1997 0.02334 0.06533 0.36 0.72
## Year1998 0.00750 0.06483 0.12 0.91
## Year1999 0.00688 0.06391 0.11 0.91
## Year2000 0.04650 0.06670 0.70 0.49
## Year2001 -0.08037 0.06071 -1.32 0.19
## Year2002 0.00952 0.05679 0.17 0.87
## Year2003 0.04388 0.05901 0.74 0.46
## Year2004 -0.04000 0.05915 -0.68 0.50
## Year2005 0.00471 0.05708 0.08 0.93
```

```

## Year2006          0.01806      0.05444      0.33      0.74
## Year2007          0.07541      0.05558      1.36      0.18
## Year2008         -0.07824      0.05758     -1.36      0.17
## Year2009          0.01502      0.06023      0.25      0.80
## Year2010         -0.04097      0.06116     -0.67      0.50
## Year2011         -0.01964      0.06203     -0.32      0.75
## Year2012          0.06266      0.05512      1.14      0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0141, Adjusted R-squared:  0.00223
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1379 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.269  0.877   0.947   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      6.58e-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 1.096 1          1.047
## Year              1.096 16          1.003

```


Residuals from first author



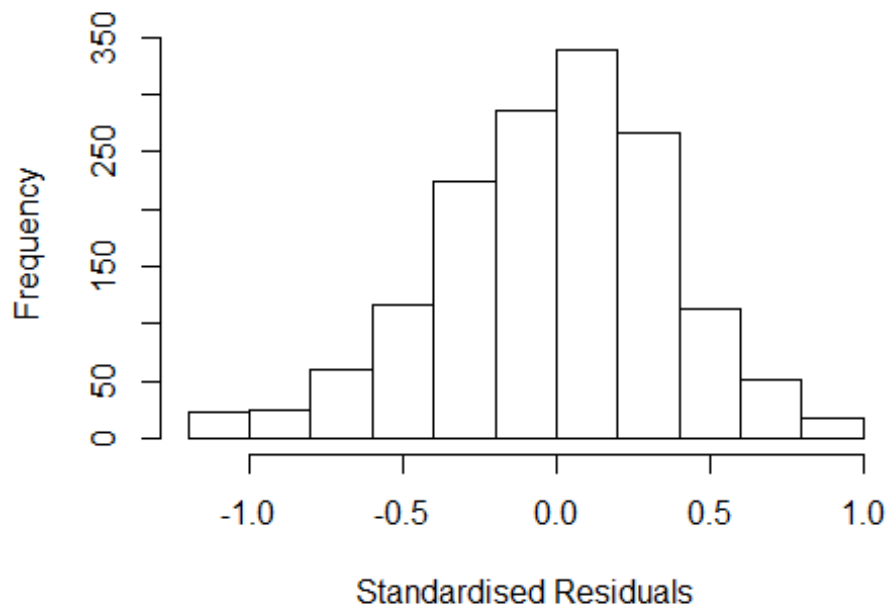
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1896 -0.2524 0.0137 0.2500 0.9810
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11424 0.04537 24.56 <2e-16 ***
## FirstAuthorFemale1 0.01555 0.02032 0.77 0.44
## Year1997 0.02208 0.06520 0.34 0.73
## Year1998 0.00540 0.06476 0.08 0.93
## Year1999 0.00650 0.06382 0.10 0.92
## Year2000 0.04684 0.06661 0.70 0.48
## Year2001 -0.08160 0.06055 -1.35 0.18
## Year2002 0.00775 0.05658 0.14 0.89
## Year2003 0.04382 0.05902 0.74 0.46
## Year2004 -0.04044 0.05904 -0.68 0.49
## Year2005 0.00461 0.05699 0.08 0.94
## Year2006 0.01581 0.05421 0.29 0.77
```

```

## Year2007          0.07410      0.05539      1.34      0.18
## Year2008         -0.07956      0.05746     -1.38      0.17
## Year2009          0.01350      0.06009      0.22      0.82
## Year2010         -0.04321      0.06108     -0.71      0.48
## Year2011         -0.02233      0.06190     -0.36      0.72
## Year2012          0.05984      0.05473      1.09      0.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00237
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1377 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.877   0.947   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      6.58e-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 1.032 1          1.016
## Year            1.032 16          1.001

```

Residuals from last author



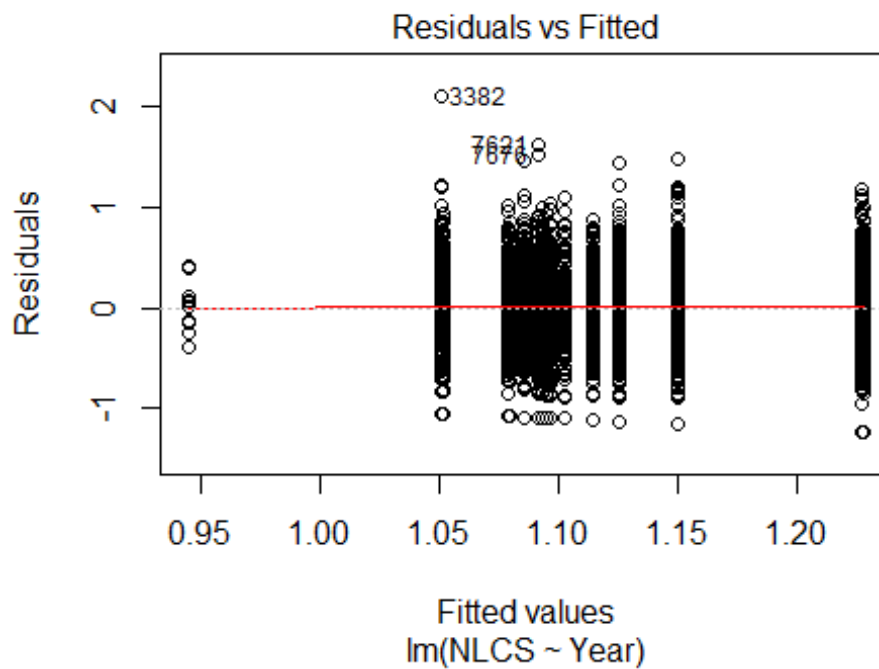
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2018 -0.2526 0.0145 0.2527 0.9676
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12454 0.04394 25.59 <2e-16 ***
## LastAuthorFemale1 -0.01770 0.02297 -0.77 0.44
## Year1997 0.02183 0.06527 0.33 0.74
## Year1998 0.00547 0.06465 0.08 0.93
## Year1999 0.00887 0.06413 0.14 0.89
## Year2000 0.04382 0.06647 0.66 0.51
## Year2001 -0.08227 0.06047 -1.36 0.17
## Year2002 0.00925 0.05682 0.16 0.87
## Year2003 0.04464 0.05917 0.75 0.45
## Year2004 -0.03833 0.05943 -0.64 0.52
## Year2005 0.00569 0.05728 0.10 0.92
## Year2006 0.01898 0.05465 0.35 0.73
```

```

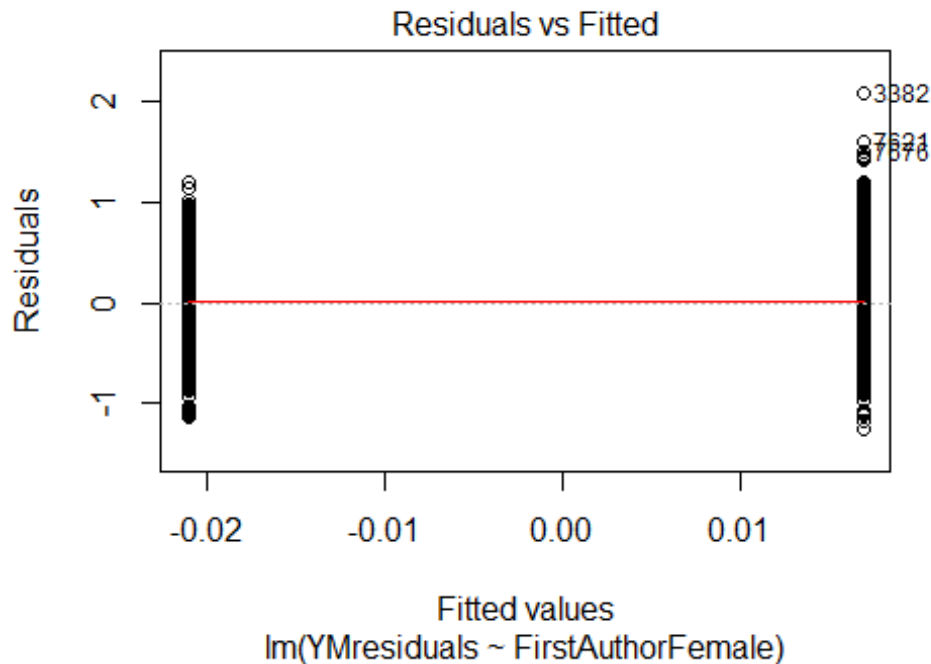
## Year2007          0.07729      0.05578      1.39      0.17
## Year2008         -0.07568      0.05767     -1.31      0.19
## Year2009          0.01657      0.06033      0.27      0.78
## Year2010         -0.03889      0.06137     -0.63      0.53
## Year2011         -0.01744      0.06202     -0.28      0.78
## Year2012          0.06506      0.05537      1.18      0.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00238
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 140 weights are ~= 1. The remaining 1380 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.261  0.877  0.947  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      6.58e-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: 1520"
## [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
##  452  464   37  384   32  442  441  370  454  414  430  420  478  506  466
## 2011 2012
##  490  518
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  203  212   21  196   11  175  234  218  281  267  263  260  294  328  289
## 2011 2012

```

```
## 325 331
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 181 194 19 167 9 157 198 181 233 243 233 228 263 295 256
## 2011 2012
## 292 300
## [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
```

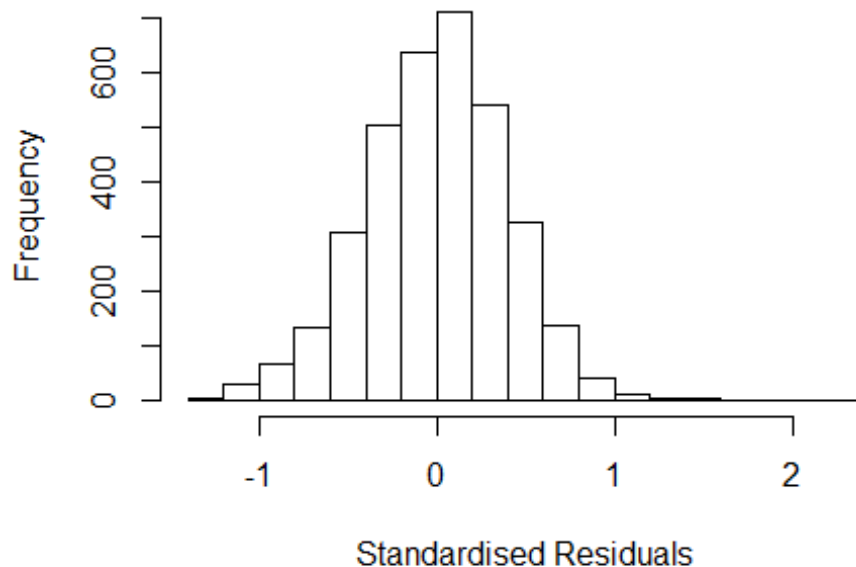


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 22, df = 1, p-value = 3e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.88629434480583"
## [1] "Male first author team size 2018 geometric mean: 4.43088101691076"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.14321511545224"
## [1] "Male last author team size 2018 geometric mean: 4.4428446452055"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, 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 1.045 1      1.022
## LastAuthorFemale  1.033 1      1.016
## UniqueAuthors    1.130 4      1.015
## Year              1.136 16     1.004
```

Residuals from first and last author and team size



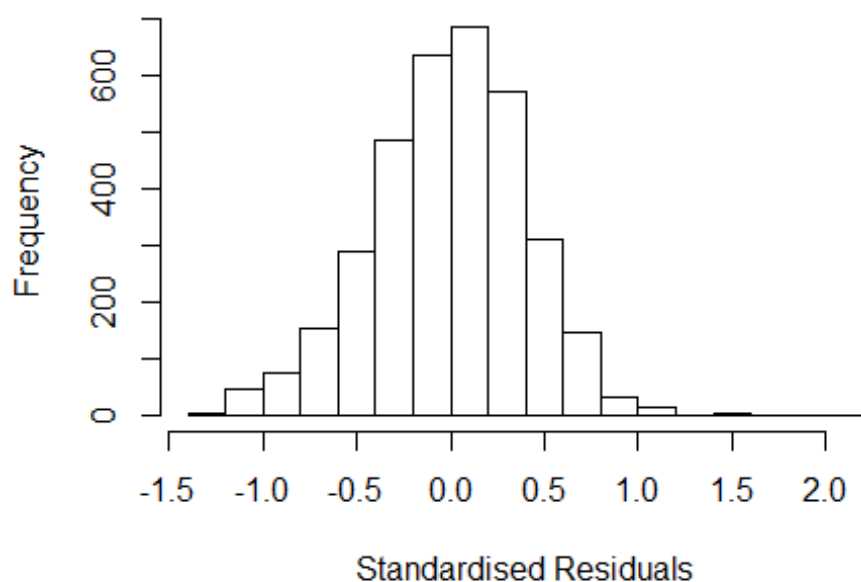
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24517 -0.26502 0.00889 0.26236 2.23412
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1164 0.0438 25.49 < 2e-16 ***
## FirstAuthorFemale1 -0.0507 0.0139 -3.64 0.00028 ***
## LastAuthorFemale1 -0.0528 0.0169 -3.13 0.00177 **
## UniqueAuthors2 0.1060 0.0363 2.92 0.00354 **
## UniqueAuthors3 0.1451 0.0349 4.16 3.2e-05 ***
## UniqueAuthors4 0.2011 0.0350 5.74 1.0e-08 ***
## UniqueAuthors5 0.2602 0.0337 7.73 1.5e-14 ***
## Year1997 -0.0163 0.0455 -0.36 0.71954
## Year1998 -0.1871 0.1053 -1.78 0.07571 .
## Year1999 -0.1511 0.0439 -3.44 0.00059 ***
```

```

## Year2000          -0.2649      0.0732   -3.62  0.00030 ***
## Year2001          -0.1311      0.0457   -2.87  0.00418 **
## Year2002          -0.1907      0.0429   -4.45  8.9e-06 ***
## Year2003          -0.1358      0.0413   -3.29  0.00102 **
## Year2004          -0.2005      0.0426   -4.71  2.6e-06 ***
## Year2005          -0.1625      0.0415   -3.92  9.2e-05 ***
## Year2006          -0.1689      0.0426   -3.96  7.6e-05 ***
## Year2007          -0.1478      0.0425   -3.48  0.00052 ***
## Year2008          -0.1705      0.0431   -3.96  7.6e-05 ***
## Year2009          -0.1365      0.0417   -3.27  0.00108 **
## Year2010          -0.1690      0.0426   -3.97  7.4e-05 ***
## Year2011          -0.1046      0.0428   -2.44  0.01455 *
## Year2012          -0.1751      0.0429   -4.08  4.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.0472
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1116 is an outlier with |weight| = 0 ( < 2.9e-05);
## 284 weights are ~= 1. The remaining 3164 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0179 0.8720 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           2.90e-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 1.035 1 1.017
## LastAuthorFemale 1.025 1 1.013
## Year 1.032 16 1.001

```


Residuals from first and last author



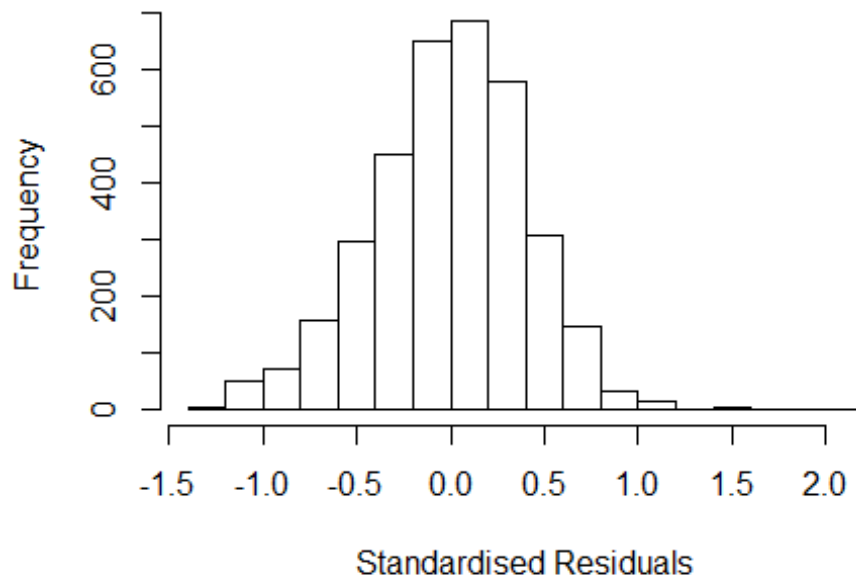
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2570 -0.2664 0.0123 0.2694 2.0634
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25699 0.03356 37.46 < 2e-16 ***
## FirstAuthorFemale1 -0.03385 0.01410 -2.40 0.01643 *
## LastAuthorFemale1 -0.05680 0.01707 -3.33 0.00088 ***
## Year1997 -0.00139 0.04506 -0.03 0.97530
## Year1998 -0.15222 0.11136 -1.37 0.17175
## Year1999 -0.13488 0.04268 -3.16 0.00159 **
## Year2000 -0.24704 0.08881 -2.78 0.00544 **
## Year2001 -0.11309 0.04518 -2.50 0.01236 *
## Year2002 -0.15828 0.04263 -3.71 0.00021 ***
## Year2003 -0.10322 0.04088 -2.53 0.01161 *
## Year2004 -0.17036 0.04246 -4.01 6.1e-05 ***
## Year2005 -0.12363 0.04062 -3.04 0.00235 **
```

```

## Year2006          -0.13074      0.04175      -3.13      0.00175 **
## Year2007          -0.10377      0.04179      -2.48      0.01307 *
## Year2008          -0.13007      0.04245      -3.06      0.00220 **
## Year2009          -0.09457      0.04172      -2.27      0.02346 *
## Year2010          -0.12999      0.04224      -3.08      0.00211 **
## Year2011          -0.06700      0.04297      -1.56      0.11903
## Year2012          -0.13348      0.04273      -3.12      0.00180 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.0139
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1116 is an outlier with |weight| = 0 ( < 2.9e-05);
## 264 weights are ~= 1. The remaining 3184 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0464 0.8750 0.9500 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          2.90e-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 1.021 1          1.011
## Year              1.021 16          1.001

```

Residuals from first author



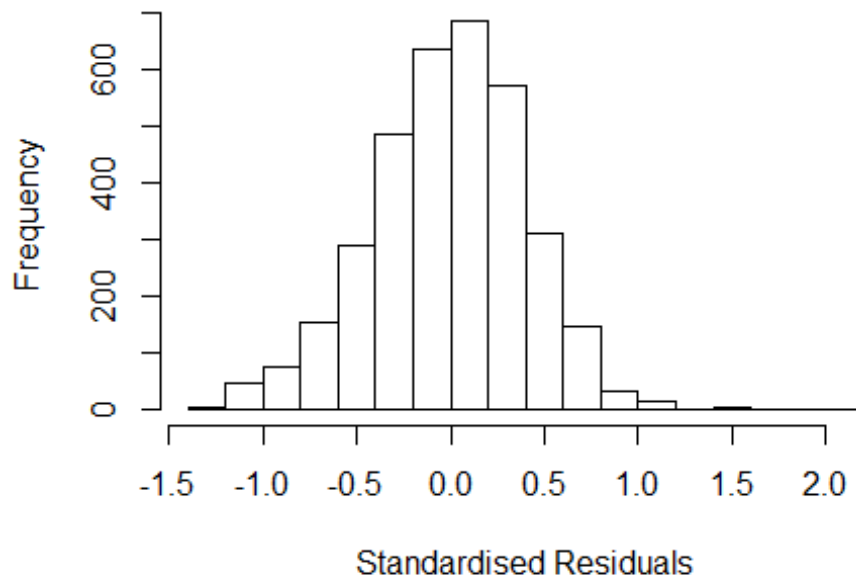
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.249 -0.265 0.014 0.269 2.074
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24936 0.03346 37.34 < 2e-16 ***
## FirstAuthorFemale1 -0.03841 0.01404 -2.74 0.00626 **
## Year1997 -0.00481 0.04502 -0.11 0.91499
## Year1998 -0.14845 0.11245 -1.32 0.18688
## Year1999 -0.13808 0.04284 -3.22 0.00128 **
## Year2000 -0.24446 0.08863 -2.76 0.00584 **
## Year2001 -0.11430 0.04536 -2.52 0.01178 *
## Year2002 -0.15704 0.04273 -3.67 0.00024 ***
## Year2003 -0.10698 0.04095 -2.61 0.00903 **
## Year2004 -0.17304 0.04233 -4.09 4.5e-05 ***
## Year2005 -0.12638 0.04059 -3.11 0.00186 **
## Year2006 -0.13537 0.04170 -3.25 0.00118 **
```

```

## Year2007          -0.10935      0.04184      -2.61  0.00900 **
## Year2008          -0.13227      0.04238      -3.12  0.00182 **
## Year2009          -0.10030      0.04168      -2.41  0.01616 *
## Year2010          -0.13567      0.04241      -3.20  0.00139 **
## Year2011          -0.07138      0.04279      -1.67  0.09536 .
## Year2012          -0.13834      0.04263      -3.25  0.00118 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0155, Adjusted R-squared:  0.0107
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1116 is an outlier with |weight| = 0 ( < 2.9e-05);
## 272 weights are ~= 1. The remaining 3176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0659 0.8740 0.9520 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          2.90e-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 1.013 1          1.006
## Year          1.013 16          1.000

```

Residuals from last author



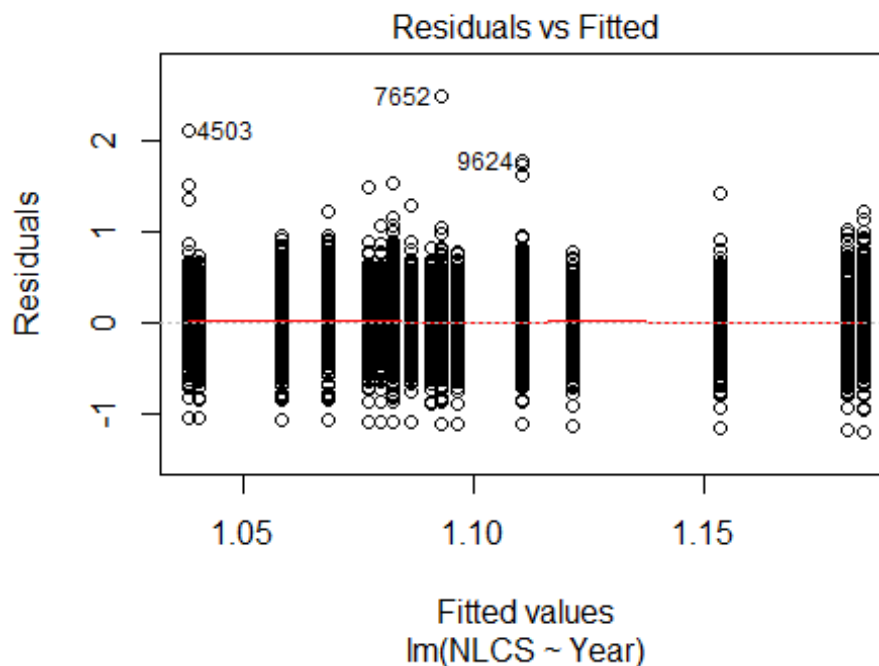
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2451 -0.2664 0.0103 0.2686 2.0784
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24512 0.03329 37.41 < 2e-16 ***
## LastAuthorFemale1 -0.06059 0.01698 -3.57 0.00036 ***
## Year1997 -0.00182 0.04521 -0.04 0.96795
## Year1998 -0.15641 0.10939 -1.43 0.15285
## Year1999 -0.13718 0.04285 -3.20 0.00138 **
## Year2000 -0.24907 0.08853 -2.81 0.00493 **
## Year2001 -0.11411 0.04530 -2.52 0.01181 *
## Year2002 -0.15884 0.04284 -3.71 0.00021 ***
## Year2003 -0.10568 0.04120 -2.57 0.01036 *
## Year2004 -0.17348 0.04262 -4.07 4.8e-05 ***
## Year2005 -0.12700 0.04079 -3.11 0.00186 **
## Year2006 -0.13429 0.04193 -3.20 0.00137 **
```

```

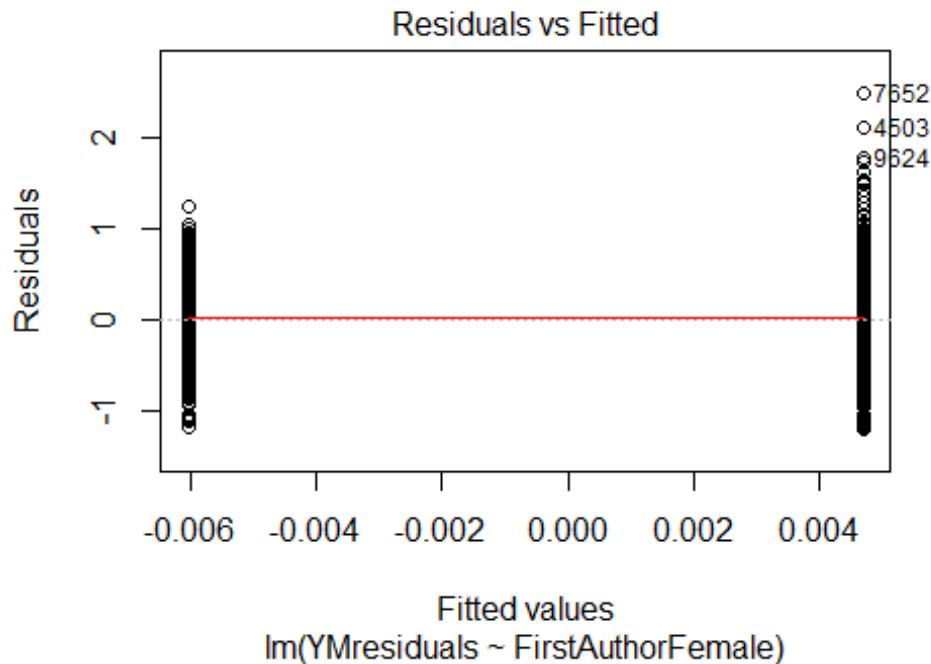
## Year2007          -0.10772      0.04201    -2.56  0.01039 *
## Year2008          -0.13412      0.04266    -3.14  0.00168 **
## Year2009          -0.09855      0.04195    -2.35  0.01887 *
## Year2010          -0.13267      0.04240    -3.13  0.00177 **
## Year2011          -0.07137      0.04312    -1.66  0.09796 .
## Year2012          -0.13898      0.04290    -3.24  0.00121 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.0124
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1116 is an outlier with |weight| = 0 ( < 2.9e-05);
## 270 weights are ~= 1. The remaining 3178 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0411 0.8750 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          2.90e-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: 3449"
## [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
## 487 474 448 398 462 436 467 356 416 407 445 430 517 574 572
## 2011 2012
## 558 556
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 289 290 261 251 225 241 307 235 320 294 324 296 353 405 395

```

```
## 2011 2012
## 405 392
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 259 261 240 226 208 217 274 200 284 269 293 264 329 357 351
## 2011 2012
## 358 356
## [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
```

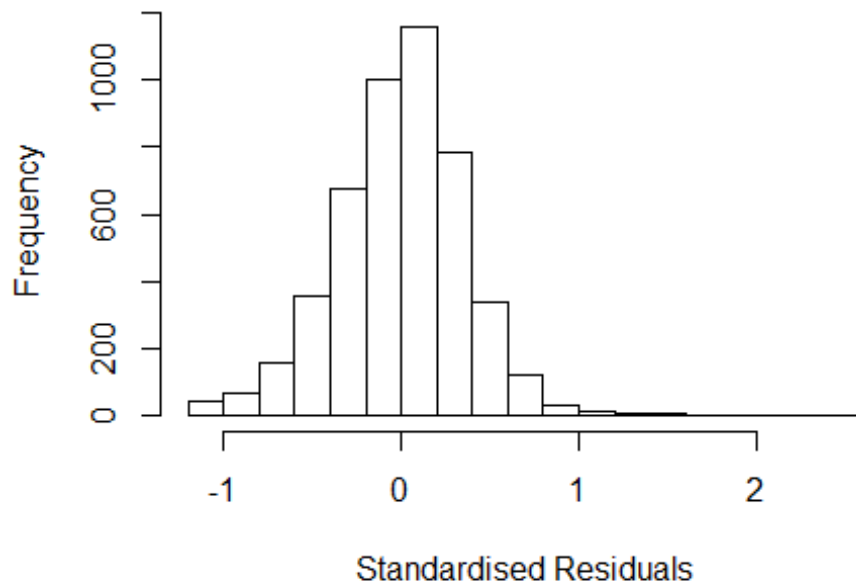


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 38, df = 1, p-value = 9e-10
```



```
## [1] "Female first author team size 2018 geometric mean: 5.26769280698074"
## [1] "Male first author team size 2018 geometric mean: 4.08282624045342"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.95892837907113"
## [1] "Male last author team size 2018 geometric mean: 4.35497388675025"
##
## 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] "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  1.029 1      1.014
## UniqueAuthors     1.147 4      1.017
## Year              1.171 16      1.005
```


Residuals from first and last author and team size



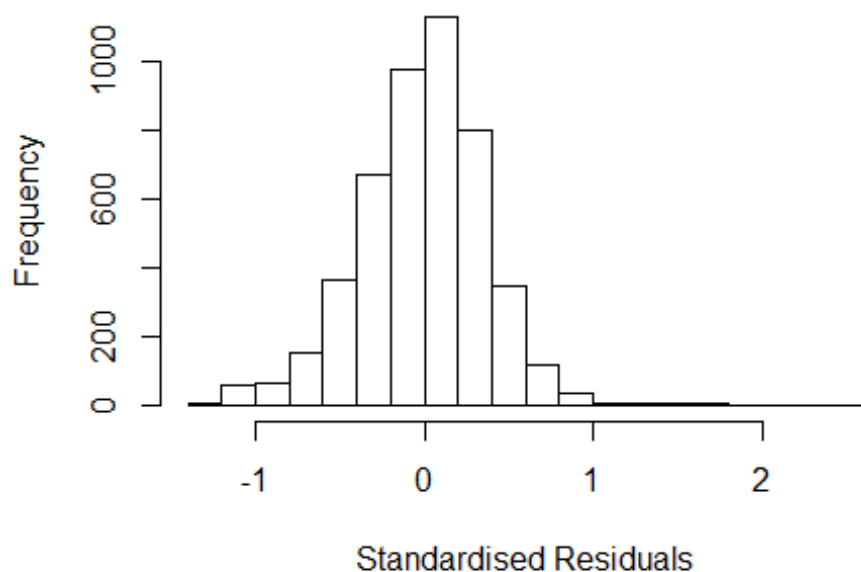
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7652 75249087100 3.574 2009      1303      6      2.518
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.194 -0.225  0.013  0.223  2.518
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0335     0.0420  24.61 < 2e-16 ***
## FirstAuthorFemale1 -0.0175     0.0102  -1.72  0.08590 .
## LastAuthorFemale1 -0.0461     0.0128  -3.60  0.00032 ***
## UniqueAuthors2      0.1368     0.0381   3.59  0.00034 ***
## UniqueAuthors3      0.1740     0.0370   4.70  2.7e-06 ***
## UniqueAuthors4      0.2025     0.0370   5.47  4.7e-08 ***
## UniqueAuthors5      0.2940     0.0362   8.13  5.6e-16 ***
## Year1997          -0.0230     0.0336  -0.68  0.49346
## Year1998          -0.0223     0.0366  -0.61  0.54273
## Year1999          -0.0948     0.0343  -2.76  0.00572 **
```

```

## Year2000          -0.1367      0.0352   -3.88  0.00010 ***
## Year2001          -0.1661      0.0348   -4.77  1.9e-06 ***
## Year2002          -0.1411      0.0325   -4.34  1.5e-05 ***
## Year2003          -0.1396      0.0326   -4.28  1.9e-05 ***
## Year2004          -0.2015      0.0321   -6.29  3.6e-10 ***
## Year2005          -0.1706      0.0321   -5.31  1.1e-07 ***
## Year2006          -0.1484      0.0325   -4.57  4.9e-06 ***
## Year2007          -0.1326      0.0324   -4.09  4.3e-05 ***
## Year2008          -0.1816      0.0324   -5.61  2.2e-08 ***
## Year2009          -0.1513      0.0316   -4.79  1.7e-06 ***
## Year2010          -0.1480      0.0320   -4.62  3.9e-06 ***
## Year2011          -0.1331      0.0322   -4.13  3.6e-05 ***
## Year2012          -0.1506      0.0330   -4.56  5.2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.0639, Adjusted R-squared:  0.0596
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 observations c(1899,1936,2041,3371,4210,4211,4267,4462)
## are outliers with |weight| = 0 ( < 2.1e-05);
## 416 weights are ~ = 1. The remaining 4322 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8670 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          2.11e-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 1.024 1          1.012
## LastAuthorFemale 1.018 1          1.009
## Year              1.036 16          1.001

```

Residuals from first and last author



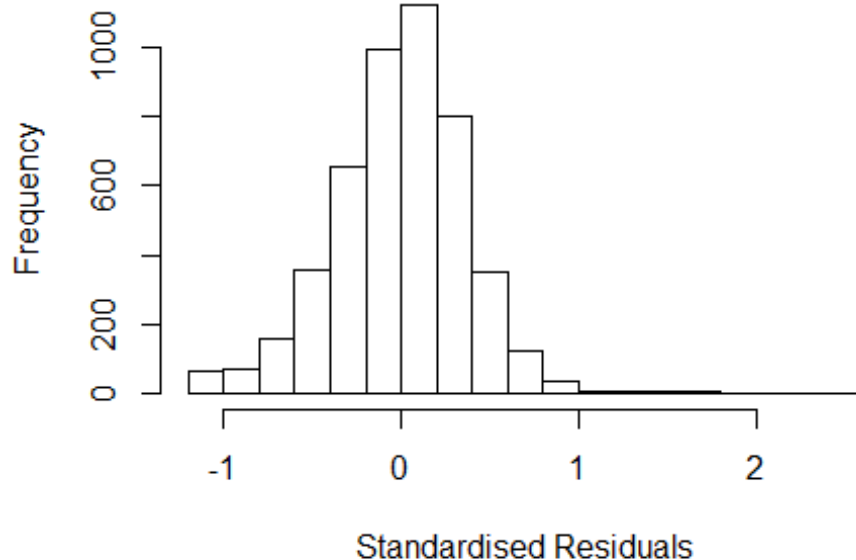
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2080 -0.2312 0.0138 0.2248 2.4661
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20801 0.02578 46.87 < 2e-16 ***
## FirstAuthorFemale1 -0.00846 0.01043 -0.81 0.41726
## LastAuthorFemale1 -0.04945 0.01307 -3.78 0.00016 ***
## Year1997 -0.00146 0.03391 -0.04 0.96561
## Year1998 -0.01666 0.03674 -0.45 0.65018
## Year1999 -0.07053 0.03376 -2.09 0.03674 *
## Year2000 -0.10491 0.03488 -3.01 0.00265 **
## Year2001 -0.13333 0.03445 -3.87 0.00011 ***
## Year2002 -0.09287 0.03258 -2.85 0.00438 **
## Year2003 -0.10266 0.03288 -3.12 0.00180 **
## Year2004 -0.15976 0.03218 -4.96 7.2e-07 ***
## Year2005 -0.12117 0.03212 -3.77 0.00016 ***
```

```

## Year2006          -0.09677    0.03205   -3.02  0.00255 **
## Year2007          -0.08057    0.03217   -2.50  0.01230 *
## Year2008          -0.13652    0.03220   -4.24  2.3e-05 ***
## Year2009          -0.10007    0.03176   -3.15  0.00164 **
## Year2010          -0.09927    0.03189   -3.11  0.00186 **
## Year2011          -0.08186    0.03209   -2.55  0.01077 *
## Year2012          -0.10259    0.03290   -3.12  0.00183 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.339
## Multiple R-squared:  0.0183, Adjusted R-squared:  0.0145
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 5 observations c(1899,3371,4210,4211,4267)
## are outliers with |weight| = 0 ( < 2.1e-05);
## 376 weights are ~ = 1. The remaining 4365 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.007  0.868  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      2.11e-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 1.021 1      1.010
## Year              1.021 16      1.001

```

Residuals from first author



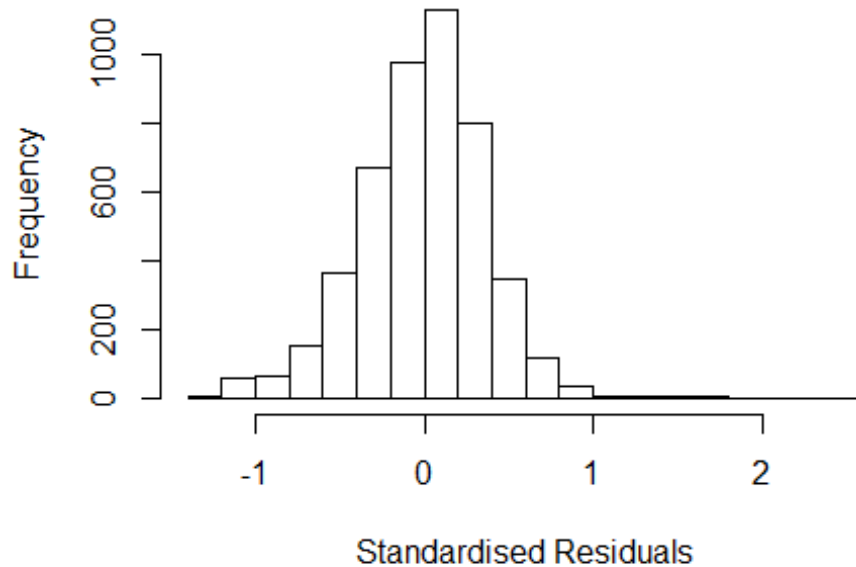
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1998 -0.2331 0.0128 0.2274 2.4758
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.199786 0.025636 46.80 < 2e-16 ***
## FirstAuthorFemale1 -0.010675 0.010441 -1.02 0.30665
## Year1997 -0.000137 0.034026 0.00 0.99679
## Year1998 -0.016087 0.036653 -0.44 0.66076
## Year1999 -0.068526 0.033652 -2.04 0.04177 *
## Year2000 -0.103484 0.034889 -2.97 0.00303 **
## Year2001 -0.131768 0.034435 -3.83 0.00013 ***
## Year2002 -0.093497 0.032493 -2.88 0.00403 **
## Year2003 -0.102320 0.032894 -3.11 0.00188 **
## Year2004 -0.159171 0.032156 -4.95 7.7e-07 ***
## Year2005 -0.120938 0.032129 -3.76 0.00017 ***
## Year2006 -0.097907 0.032071 -3.05 0.00228 **
```

```

## Year2007          -0.081545    0.032269    -2.53    0.01153 *
## Year2008          -0.136834    0.032207    -4.25    2.2e-05 ***
## Year2009          -0.101589    0.031749    -3.20    0.00138 **
## Year2010          -0.101225    0.031914    -3.17    0.00152 **
## Year2011          -0.084989    0.032062    -2.65    0.00806 **
## Year2012          -0.105260    0.032909    -3.20    0.00139 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.0115
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 5 observations c(1899,3371,4210,4211,4267)
## are outliers with |weight| = 0 ( < 2.1e-05);
## 395 weights are ~ = 1. The remaining 4346 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0055 0.8680 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          2.11e-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 1.016 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
## -1.2047 -0.2326 0.0125 0.2232 2.4698
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20468 0.02529 47.63 < 2e-16 ***
## LastAuthorFemale1 -0.05005 0.01306 -3.83 0.00013 ***
## Year1997 -0.00166 0.03391 -0.05 0.96089
## Year1998 -0.01692 0.03673 -0.46 0.64514
## Year1999 -0.07014 0.03374 -2.08 0.03767 *
## Year2000 -0.10524 0.03488 -3.02 0.00256 **
## Year2001 -0.13355 0.03446 -3.88 0.00011 ***
## Year2002 -0.09327 0.03259 -2.86 0.00423 **
## Year2003 -0.10356 0.03286 -3.15 0.00164 **
## Year2004 -0.16060 0.03215 -4.99 6.1e-07 ***
## Year2005 -0.12129 0.03210 -3.78 0.00016 ***
## Year2006 -0.09699 0.03206 -3.02 0.00250 **
```

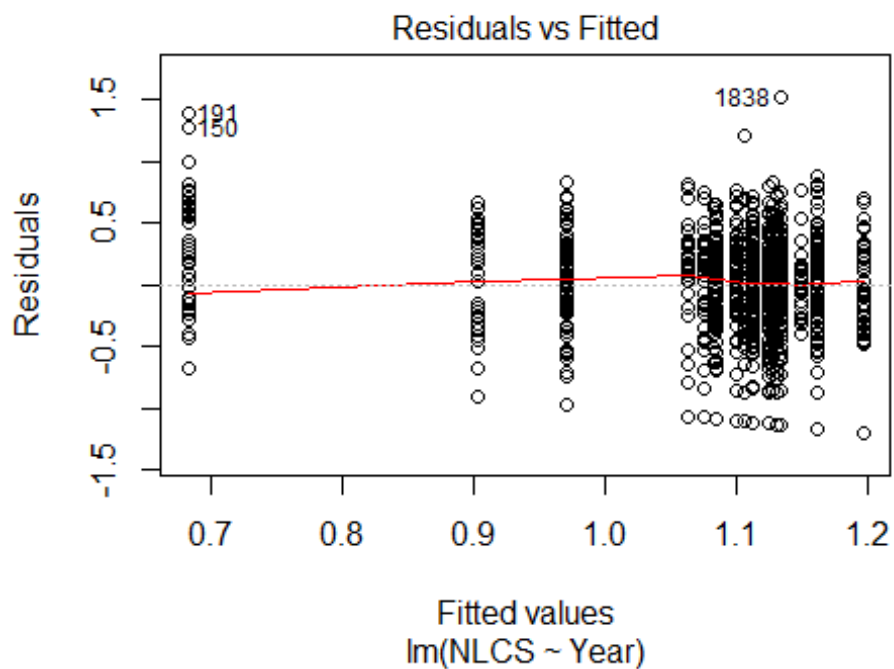
```

## Year2007          -0.08076      0.03218      -2.51   0.01212  *
## Year2008          -0.13663      0.03219      -4.24   2.2e-05  ***
## Year2009          -0.10046      0.03175      -3.16   0.00157  **
## Year2010          -0.09973      0.03187      -3.13   0.00176  **
## Year2011          -0.08271      0.03203      -2.58   0.00986  **
## Year2012          -0.10349      0.03289      -3.15   0.00166  **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.339
## Multiple R-squared:  0.0181, Adjusted R-squared:  0.0146
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 5 observations c(1899,3371,4210,4211,4267)
## are outliers with |weight| = 0 ( < 2.1e-05);
## 379 weights are ~ = 1. The remaining 4362 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0061 0.8680 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          2.11e-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: 4746"
## [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
##   41   51   74   61   76   75   65   56   95   80  112  131  147  165  161
## 2011 2012
##   176  159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

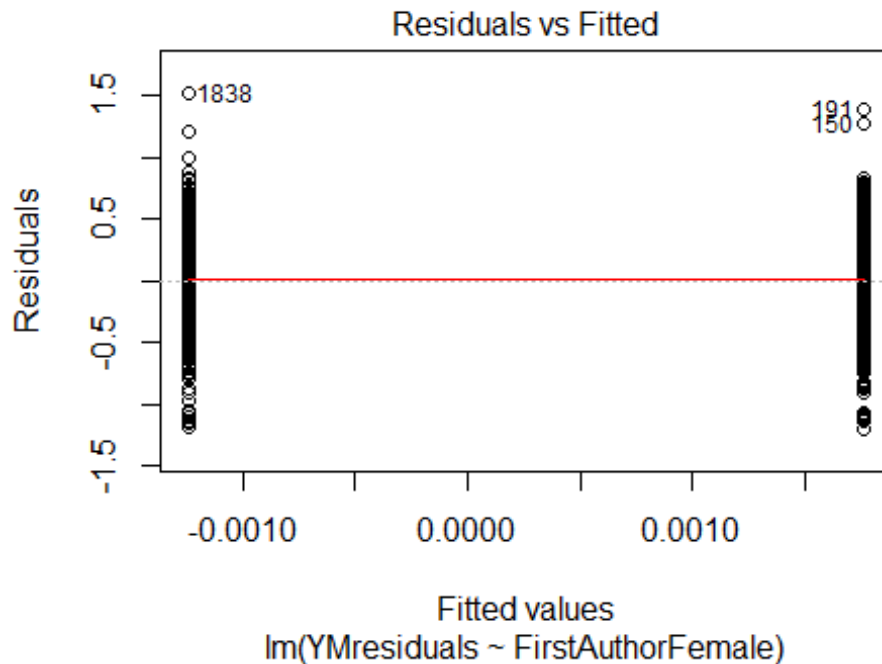
```



```
##      27      36      57      47      47      40      46      41      67      66      79      92     100     116     103
## 2011 2012
##    134     95
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    24    33    55    46    43    35    40    39    64    62    74    83    93   101    94
## 2011 2012
##    119     84
## [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
```

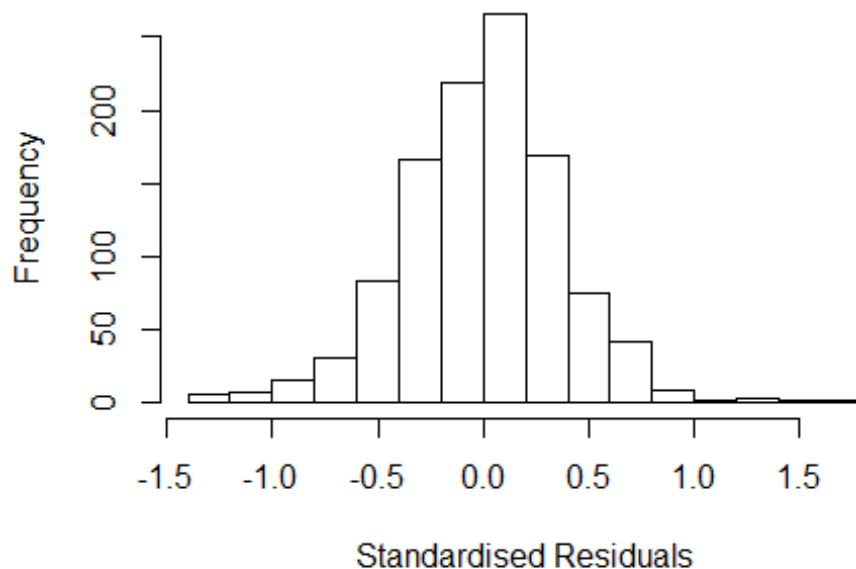


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.22, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 5.81592625975823"
## [1] "Male first author team size 2018 geometric mean: 5.74250558264951"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.66883773930689"
## [1] "Male last author team size 2018 geometric mean: 6.26120858791008"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, 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 1.064 1      1.032
## LastAuthorFemale  1.130 1      1.063
## UniqueAuthors    1.440 4      1.047
## Year             1.591 16      1.015
```

Residuals from first and last author and team size



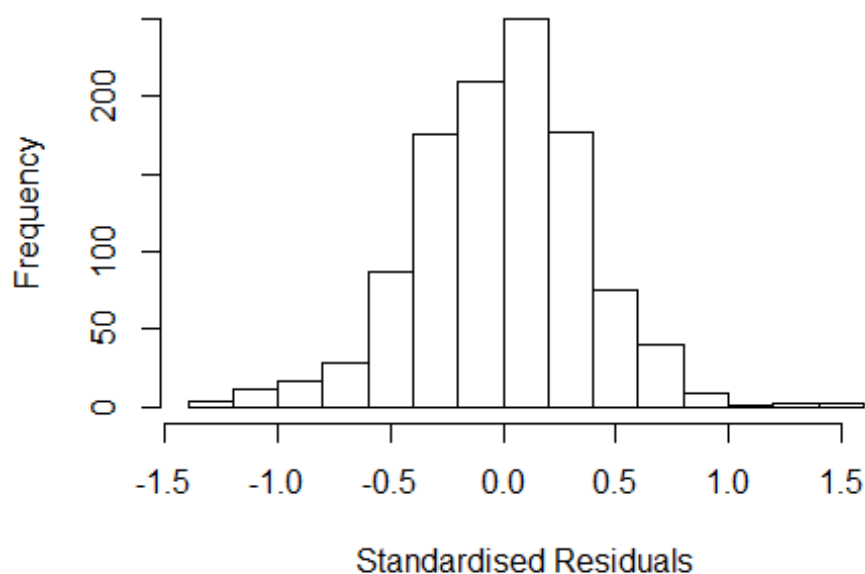
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3078 -0.2339 0.0109 0.2324 1.6410
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8831 0.0937 9.43 < 2e-16 ***
## FirstAuthorFemale1 -0.0030 0.0226 -0.13 0.89435
## LastAuthorFemale1 0.0282 0.0282 1.00 0.31720
## UniqueAuthors2 0.1948 0.0862 2.26 0.02397 *
## UniqueAuthors3 0.3012 0.0817 3.69 0.00024 ***
## UniqueAuthors4 0.3190 0.0828 3.85 0.00012 ***
## UniqueAuthors5 0.3971 0.0792 5.01 6.3e-07 ***
## Year1997 0.0276 0.0877 0.31 0.75320
## Year1998 -0.5716 0.1226 -4.66 3.5e-06 ***
## Year1999 -0.0921 0.0830 -1.11 0.26783
```

```

## Year2000          0.0319      0.0854      0.37  0.70862
## Year2001         -0.1102      0.0832     -1.32  0.18547
## Year2002         -0.0552      0.0853     -0.65  0.51728
## Year2003         -0.2565      0.0950     -2.70  0.00702 **
## Year2004         -0.0828      0.0776     -1.07  0.28622
## Year2005         -0.2067      0.0799     -2.59  0.00979 **
## Year2006         -0.0981      0.0758     -1.30  0.19559
## Year2007         -0.0793      0.0730     -1.09  0.27762
## Year2008         -0.0707      0.0715     -0.99  0.32262
## Year2009         -0.1012      0.0678     -1.49  0.13574
## Year2010         -0.0941      0.0701     -1.34  0.17964
## Year2011         -0.1145      0.0674     -1.70  0.08954 .
## Year2012         -0.0822      0.0750     -1.10  0.27354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.333
## Multiple R-squared:  0.175, Adjusted R-squared:  0.158
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 2 observations c(109,860) are outliers with |weight| = 0 ( < 9.2e-05);
## 93 weights are ~= 1. The remaining 994 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0747 0.8590 0.9470 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      9.18e-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 1.049 1      1.024
## LastAuthorFemale  1.110 1      1.053
## Year              1.144 16      1.004

```

Residuals from first and last author



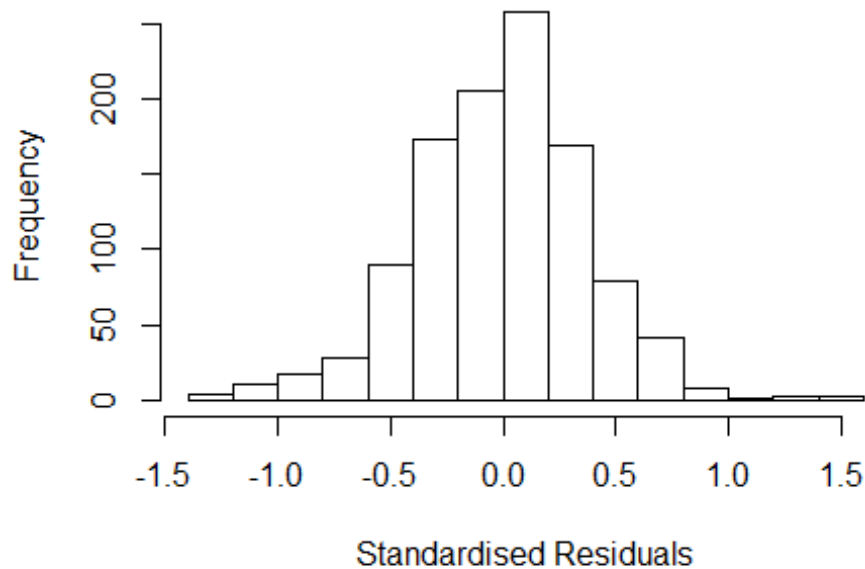
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2441 -0.2462  0.0117  0.2356  1.5131
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.13623    0.05657   20.09  <2e-16 ***
## FirstAuthorFemale1  0.00144    0.02303    0.06   0.950
## LastAuthorFemale1  0.01847    0.02941    0.63   0.530
## Year1997          0.07971    0.09103    0.88   0.381
## Year1998         -0.55065    0.14006   -3.93   9e-05 ***
## Year1999         -0.03374    0.07974   -0.42   0.672
## Year2000          0.10786    0.07922    1.36   0.174
## Year2001         -0.05885    0.08014   -0.73   0.463
## Year2002          0.00381    0.08423    0.05   0.964
## Year2003         -0.19298    0.09779   -1.97   0.049 *
## Year2004          0.00497    0.07294    0.07   0.946
## Year2005         -0.15928    0.07587   -2.10   0.036 *
```

```

## Year2006      -0.01049    0.06937   -0.15    0.880
## Year2007      0.01572    0.06793    0.23    0.817
## Year2008      0.00190    0.06600    0.03    0.977
## Year2009     -0.01093    0.06243   -0.18    0.861
## Year2010     -0.01482    0.06483   -0.23    0.819
## Year2011     -0.02053    0.06171   -0.33    0.739
## Year2012      0.02059    0.06942    0.30    0.767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0929
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 1005 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0152 0.8690 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      9.18e-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 1.041 1      1.020
## Year      1.041 16      1.001

```

Residuals from first author



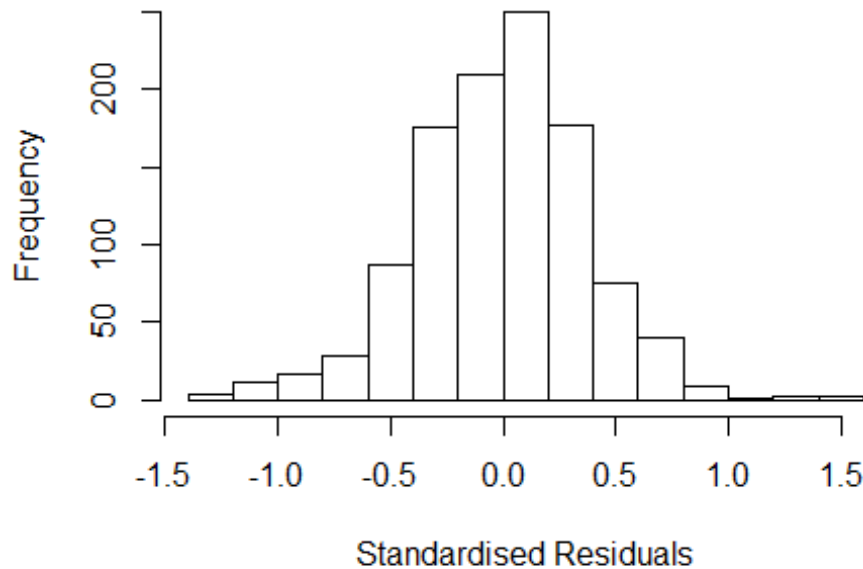
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2455 -0.2457 0.0134 0.2333 1.5294
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14014 0.05571 20.47 < 2e-16 ***
## FirstAuthorFemale1 0.00205 0.02313 0.09 0.930
## Year1997 0.07956 0.09056 0.88 0.380
## Year1998 -0.55285 0.14142 -3.91 9.8e-05 ***
## Year1999 -0.03526 0.07964 -0.44 0.658
## Year2000 0.10535 0.07905 1.33 0.183
## Year2001 -0.06070 0.07991 -0.76 0.448
## Year2002 0.00585 0.08413 0.07 0.945
## Year2003 -0.19379 0.09743 -1.99 0.047 *
## Year2004 0.00430 0.07283 0.06 0.953
## Year2005 -0.16146 0.07521 -2.15 0.032 *
## Year2006 -0.01331 0.06883 -0.19 0.847
```

```

## Year2007          0.01560      0.06784      0.23      0.818
## Year2008          0.00113      0.06588      0.02      0.986
## Year2009         -0.01067      0.06229     -0.17      0.864
## Year2010         -0.01658      0.06446     -0.26      0.797
## Year2011         -0.01962      0.06164     -0.32      0.750
## Year2012          0.02127      0.06918      0.31      0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.344
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0934
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 1000 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0104 0.8680 0.9460 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      9.18e-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 1.097 1      1.048
## Year              1.097 16      1.003

```


Residuals from last author



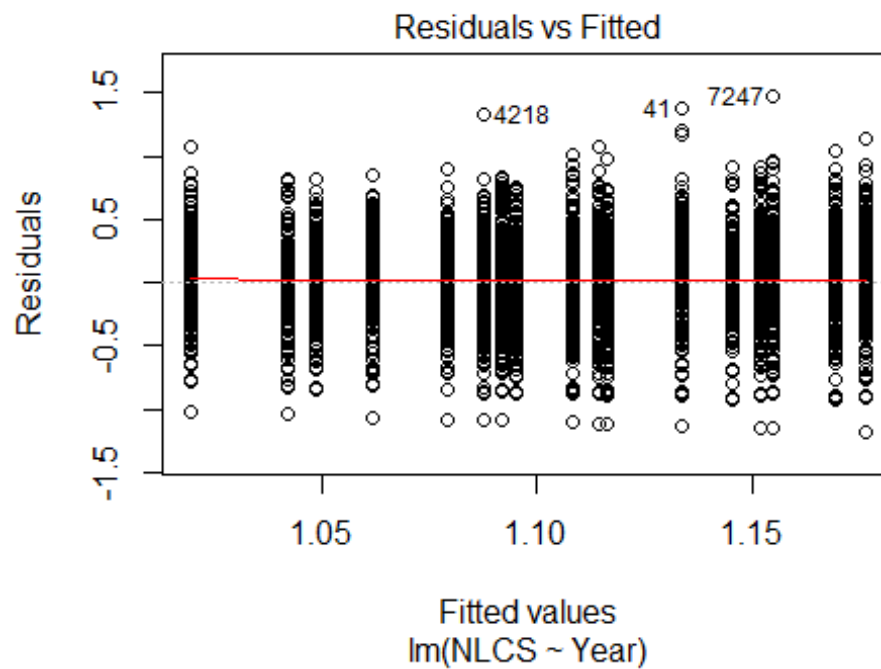
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2447 -0.2459 0.0119 0.2349 1.5123
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13684 0.05580 20.37 < 2e-16 ***
## LastAuthorFemale1 0.01859 0.02953 0.63 0.529
## Year1997 0.08005 0.09100 0.88 0.379
## Year1998 -0.55144 0.13991 -3.94 8.6e-05 ***
## Year1999 -0.03393 0.07983 -0.43 0.671
## Year2000 0.10785 0.07929 1.36 0.174
## Year2001 -0.05891 0.08022 -0.73 0.463
## Year2002 0.00369 0.08402 0.04 0.965
## Year2003 -0.19293 0.09787 -1.97 0.049 *
## Year2004 0.00509 0.07300 0.07 0.944
## Year2005 -0.15930 0.07597 -2.10 0.036 *
## Year2006 -0.01030 0.06941 -0.15 0.882
```

```

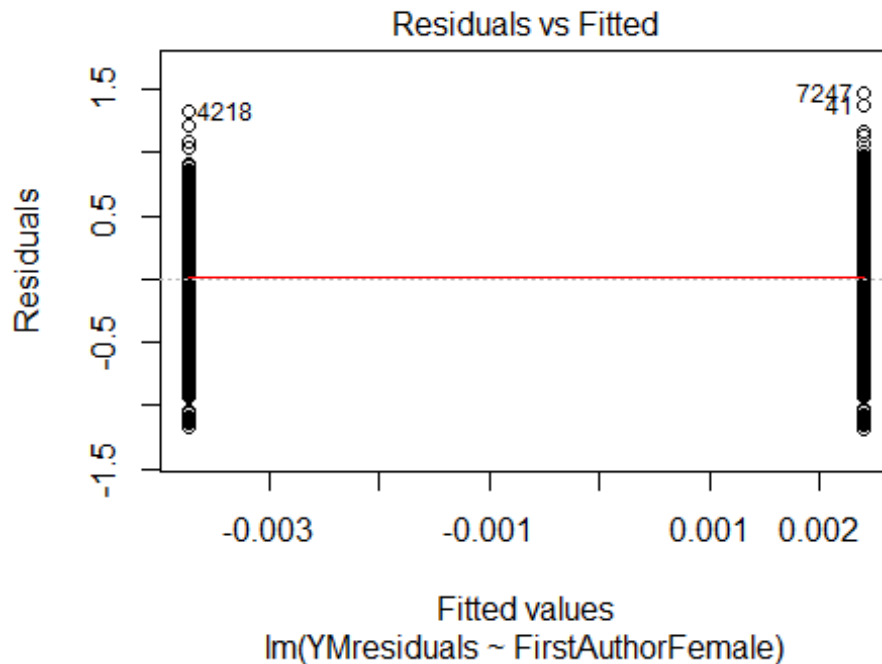
## Year2007          0.01565      0.06792      0.23      0.818
## Year2008          0.00189      0.06606      0.03      0.977
## Year2009         -0.01092      0.06249     -0.17      0.861
## Year2010         -0.01472      0.06490     -0.23      0.821
## Year2011         -0.02055      0.06177     -0.33      0.739
## Year2012          0.02045      0.06947      0.29      0.768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0939
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 1005 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0149 0.8690 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      9.18e-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: 1089"
## [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
## 391 427 405 395 391 328 328 314 312 318 352 393 340 339 332
## 2011 2012
## 357 364
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 189 198 196 196 152 130 207 177 194 225 245 272 234 244 221
## 2011 2012

```

```
## 255 257
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 168 168 178 178 130 114 180 147 170 206 221 240 210 215 199
## 2011 2012
## 225 241
## [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
```

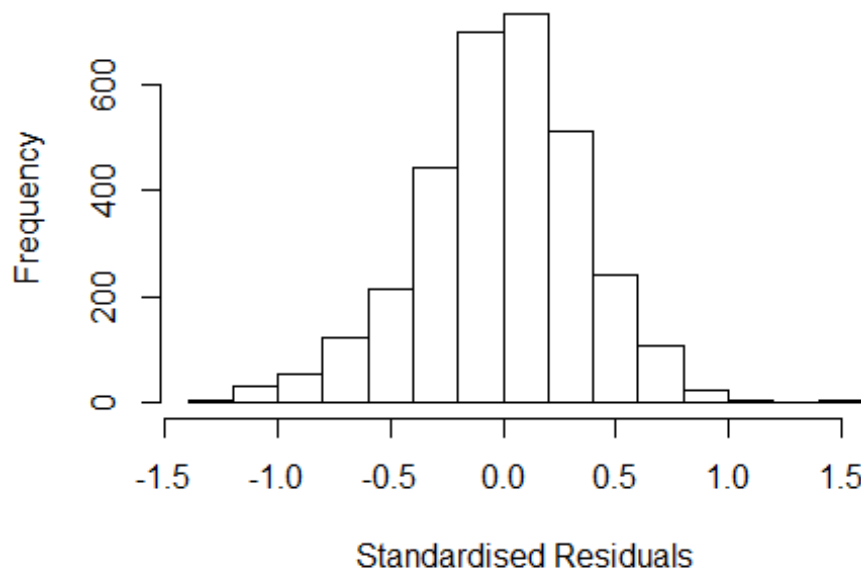


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.7, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 4.36941880706434"
## [1] "Male first author team size 2018 geometric mean: 3.99757401553917"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.28478818091243"
## [1] "Male last author team size 2018 geometric mean: 4.09497413002083"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6300, 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 1.049 1      1.024
## LastAuthorFemale  1.053 1      1.026
## UniqueAuthors     1.149 4      1.018
## Year              1.186 16      1.005
```

Residuals from first and last author and team size



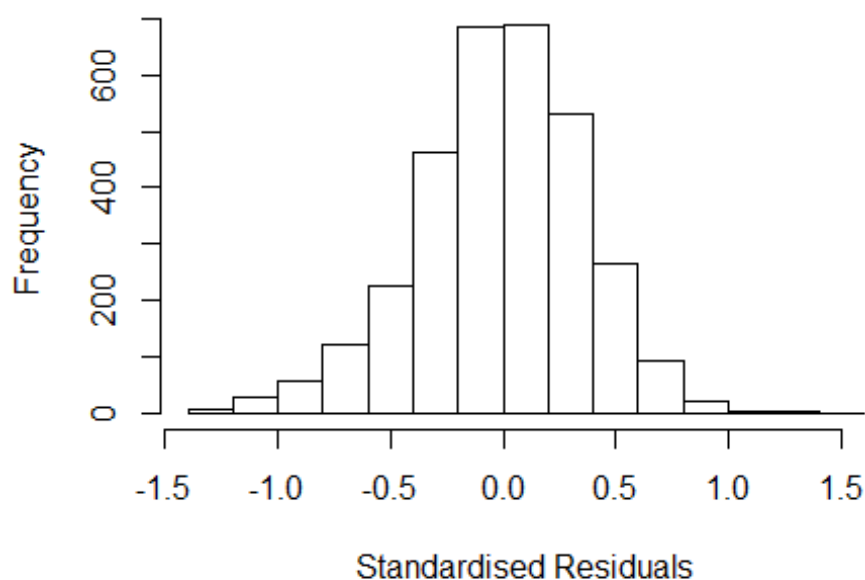
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27926 -0.22918  0.00726  0.23070  1.52152
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09020    0.04913   22.19 < 2e-16 ***
## FirstAuthorFemale1 -0.01522    0.01320   -1.15  0.24896
## LastAuthorFemale1 -0.05247    0.01639   -3.20  0.00138 **
## UniqueAuthors2     0.04279    0.03887    1.10  0.27103
## UniqueAuthors3     0.06805    0.03785    1.80  0.07230 .
## UniqueAuthors4     0.13388    0.03914    3.42  0.00063 ***
## UniqueAuthors5     0.17248    0.03741    4.61  4.2e-06 ***
## Year1997          -0.00903    0.04894   -0.18  0.85366
## Year1998          -0.11963    0.04641   -2.58  0.01000 **
## Year1999          -0.07790    0.04340   -1.79  0.07278 .
```

```

## Year2000      -0.03033    0.04488   -0.68   0.49927
## Year2001      -0.11022    0.04594   -2.40   0.01648 *
## Year2002      -0.06930    0.04166   -1.66   0.09630 .
## Year2003      -0.08842    0.04081   -2.17   0.03035 *
## Year2004      -0.12277    0.04178   -2.94   0.00332 **
## Year2005      -0.09163    0.04109   -2.23   0.02582 *
## Year2006      -0.03820    0.04053   -0.94   0.34610
## Year2007      -0.06150    0.03986   -1.54   0.12293
## Year2008      -0.07365    0.04267   -1.73   0.08446 .
## Year2009      -0.00408    0.04053   -0.10   0.91981
## Year2010      -0.03787    0.04267   -0.89   0.37486
## Year2011       0.01658    0.04175    0.40   0.69138
## Year2012      -0.02951    0.04054   -0.73   0.46675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0443, Adjusted R-squared:  0.0376
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 270 weights are ~= 1. The remaining 2920 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0085 0.8650 0.9510 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          3.13e-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 1.044 1 1.022
## LastAuthorFemale 1.043 1 1.021
## Year 1.057 16 1.002

```

Residuals from first and last author



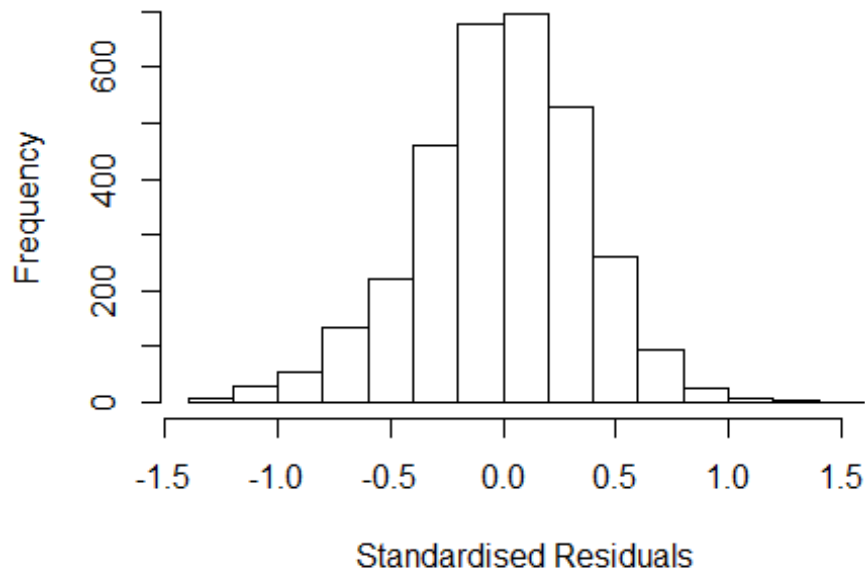
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22345 -0.23110  0.00325  0.23256  1.45306
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18166    0.03312   35.67  <2e-16 ***
## FirstAuthorFemale1 -0.00624    0.01337   -0.47   0.6408
## LastAuthorFemale1 -0.05060    0.01657   -3.05   0.0023 **
## Year1997        -0.02124    0.04852   -0.44   0.6616
## Year1998        -0.13498    0.04648   -2.90   0.0037 **
## Year1999        -0.08398    0.04316   -1.95   0.0518 .
## Year2000        -0.01767    0.04445   -0.40   0.6910
## Year2001        -0.10700    0.04523   -2.37   0.0180 *
## Year2002        -0.06899    0.04130   -1.67   0.0949 .
## Year2003        -0.08768    0.04128   -2.12   0.0338 *
## Year2004        -0.11552    0.04137   -2.79   0.0053 **
## Year2005        -0.07420    0.04096   -1.81   0.0702 .
```

```

## Year2006      -0.03136    0.04064   -0.77    0.4404
## Year2007      -0.04238    0.03997   -1.06    0.2890
## Year2008      -0.05466    0.04276   -1.28    0.2012
## Year2009       0.02006    0.04026    0.50    0.6183
## Year2010      -0.01657    0.04297   -0.39    0.6998
## Year2011       0.04179    0.04151    1.01    0.3141
## Year2012      -0.00971    0.04032   -0.24    0.8097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.0149
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 268 weights are ~= 1. The remaining 2922 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0399 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      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 1.031 1      1.015
## Year      1.031 16      1.001

```


Residuals from first author



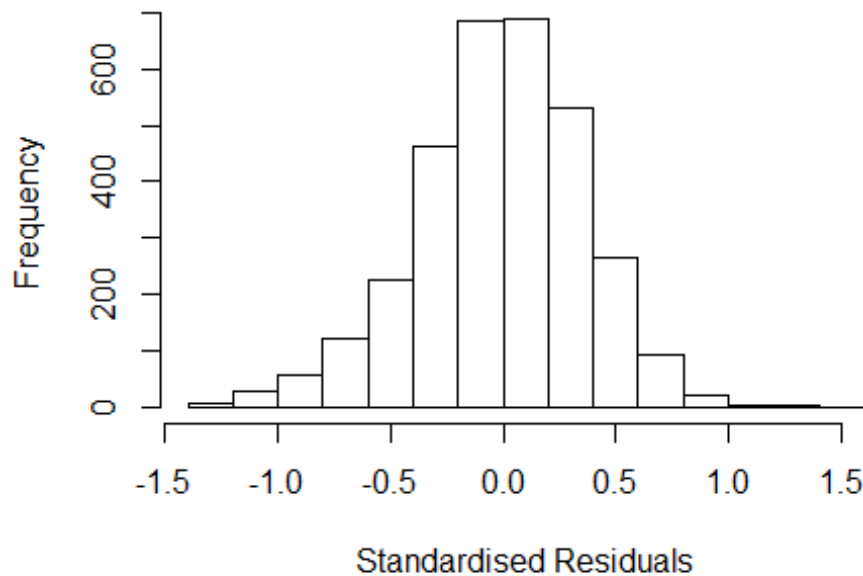
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21245 -0.23280 0.00569 0.23287 1.46280
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1746 0.0330 35.56 <2e-16 ***
## FirstAuthorFemale1 -0.0109 0.0133 -0.82 0.4140
## Year1997 -0.0227 0.0488 -0.47 0.6419
## Year1998 -0.1368 0.0465 -2.95 0.0032 **
## Year1999 -0.0872 0.0429 -2.03 0.0422 *
## Year2000 -0.0188 0.0446 -0.42 0.6737
## Year2001 -0.1103 0.0450 -2.45 0.0144 *
## Year2002 -0.0689 0.0413 -1.67 0.0952 .
## Year2003 -0.0885 0.0412 -2.15 0.0317 *
## Year2004 -0.1148 0.0413 -2.78 0.0054 **
## Year2005 -0.0747 0.0409 -1.83 0.0680 .
## Year2006 -0.0338 0.0405 -0.83 0.4045
```

```

## Year2007          -0.0436      0.0398   -1.10    0.2733
## Year2008          -0.0580      0.0426   -1.36    0.1740
## Year2009           0.0179      0.0402    0.45    0.6557
## Year2010          -0.0195      0.0429   -0.45    0.6504
## Year2011           0.0379      0.0414    0.91    0.3609
## Year2012          -0.0124      0.0402   -0.31    0.7582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared:  0.0171, Adjusted R-squared:  0.0118
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 274 weights are ~= 1. The remaining 2916 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0371 0.8670 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      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 1.029 1          1.014
## Year            1.029 16          1.001

```

Residuals from last author



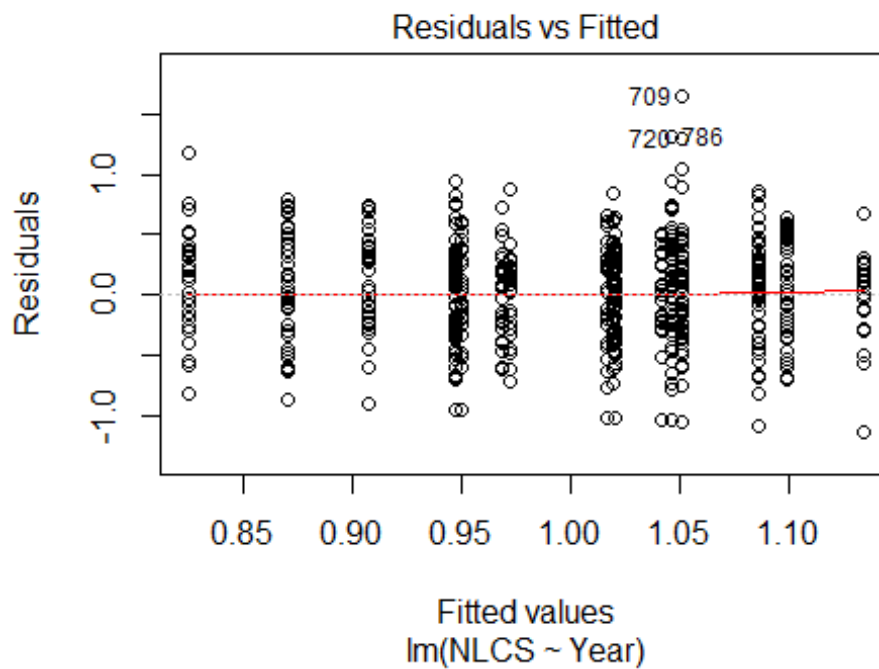
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22076 -0.23125 0.00266 0.23106 1.45585
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1799 0.0328 35.94 <2e-16 ***
## LastAuthorFemale1 -0.0514 0.0165 -3.12 0.0018 **
## Year1997 -0.0211 0.0485 -0.43 0.6644
## Year1998 -0.1356 0.0465 -2.92 0.0036 **
## Year1999 -0.0843 0.0431 -1.95 0.0507 .
## Year2000 -0.0181 0.0444 -0.41 0.6830
## Year2001 -0.1075 0.0452 -2.38 0.0175 *
## Year2002 -0.0697 0.0413 -1.69 0.0914 .
## Year2003 -0.0880 0.0413 -2.13 0.0331 *
## Year2004 -0.1162 0.0414 -2.81 0.0050 **
## Year2005 -0.0745 0.0410 -1.82 0.0693 .
## Year2006 -0.0319 0.0406 -0.79 0.4320
```

```

## Year2007          -0.0432      0.0400    -1.08    0.2797
## Year2008          -0.0549      0.0427    -1.28    0.1989
## Year2009           0.0195      0.0402     0.49    0.6272
## Year2010          -0.0171      0.0430    -0.40    0.6901
## Year2011           0.0409      0.0414     0.99    0.3238
## Year2012          -0.0107      0.0403    -0.27    0.7900
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared:  0.0204, Adjusted R-squared:  0.0151
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 271 weights are ~= 1. The remaining 2919 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0384 0.8640 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      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: 3190"
## [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
##   37   44   40   52   43   52   42   49   56   53   75   73   66   67   44
## 2011 2012
##   62   56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   24   28   35   24   26   29   36   46   43   56   64   51   47   31
## 2011 2012

```

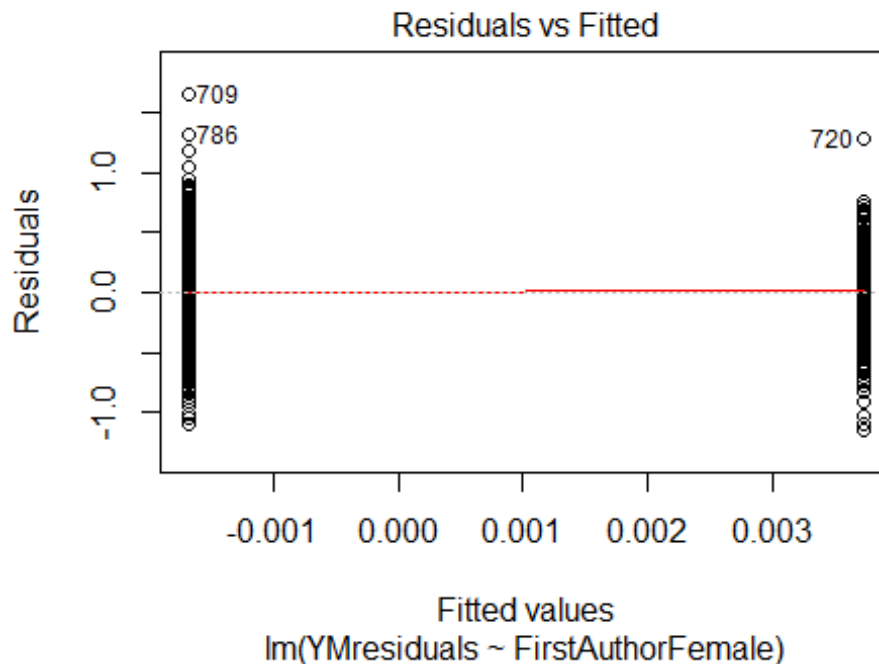
```
## 45 37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 21 26 34 24 23 23 31 42 39 52 60 51 40 24
## 2011 2012
## 35 34
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```



```
##
## 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: 6.03021022850983"
## [1] "Male first author team size 2018 geometric mean: 3.82889921219076"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

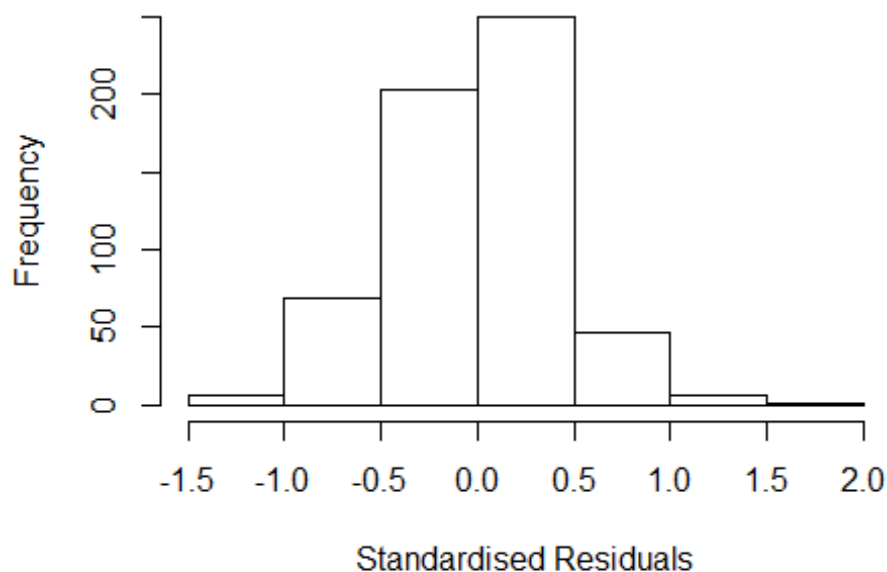
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.72594081833981"
## [1] "Male last author team size 2018 geometric mean: 4.62994965661973"

## 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 = 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 1.224  1      1.106
## LastAuthorFemale  1.132  1      1.064
## UniqueAuthors    1.628  4      1.063
## Year              1.880 16      1.020
```

Residuals from first and last author and team size



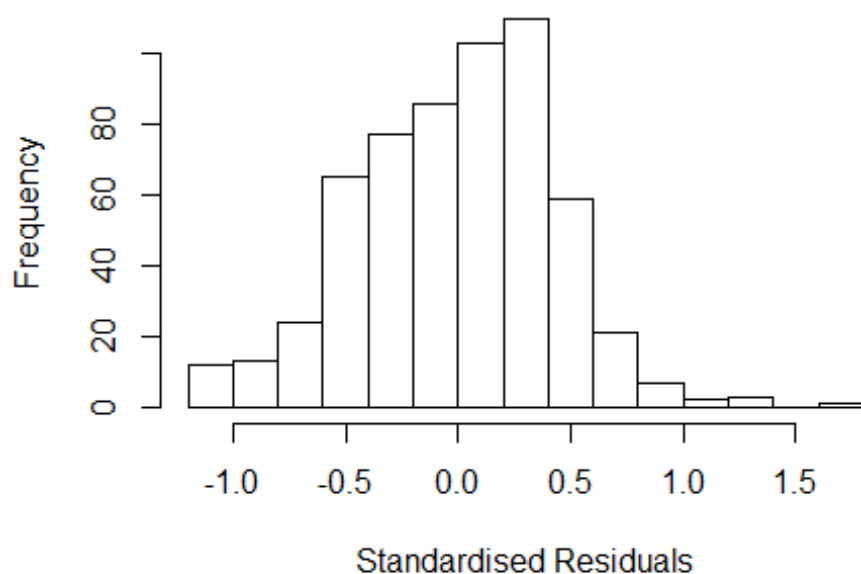
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1785 -0.2810  0.0321  0.2865  1.9939
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.70754    0.11837   5.98   4e-09 ***
## FirstAuthorFemale1 -0.01528    0.03894  -0.39  0.69486
## LastAuthorFemale1 -0.00913    0.04840  -0.19  0.85037
## UniqueAuthors2    0.30389    0.10860   2.80  0.00532 **
## UniqueAuthors3    0.30771    0.10404   2.96  0.00323 **
## UniqueAuthors4    0.22219    0.10609   2.09  0.03669 *
## UniqueAuthors5    0.38679    0.10211   3.79  0.00017 ***
## Year1997          0.15690    0.10297   1.52  0.12814
## Year1998           0.05275    0.11198   0.47  0.63775
## Year1999          -0.05676    0.09526  -0.60  0.55153
```

```

## Year2000      0.06513      0.10327      0.63      0.52848
## Year2001     -0.04874      0.12151     -0.40      0.68847
## Year2002     -0.03843      0.10626     -0.36      0.71772
## Year2003     -0.22003      0.11542     -1.91      0.05713 .
## Year2004     -0.13405      0.11745     -1.14      0.25424
## Year2005      0.05392      0.09647      0.56      0.57646
## Year2006     -0.00447      0.10309     -0.04      0.96541
## Year2007      0.08412      0.10160      0.83      0.40805
## Year2008      0.08231      0.09935      0.83      0.40775
## Year2009      0.02446      0.10328      0.24      0.81289
## Year2010     -0.16086      0.13556     -1.19      0.23586
## Year2011     -0.01752      0.10215     -0.17      0.86391
## Year2012     -0.11655      0.11191     -1.04      0.29811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.0865, Adjusted R-squared:  0.0506
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 327 is an outlier with |weight| = 0 ( < 0.00017);
## 41 weights are ~= 1. The remaining 541 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.281  0.881  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           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 1.104 1 1.051
## LastAuthorFemale 1.083 1 1.041
## Year 1.162 16 1.005

```


Residuals from first and last author



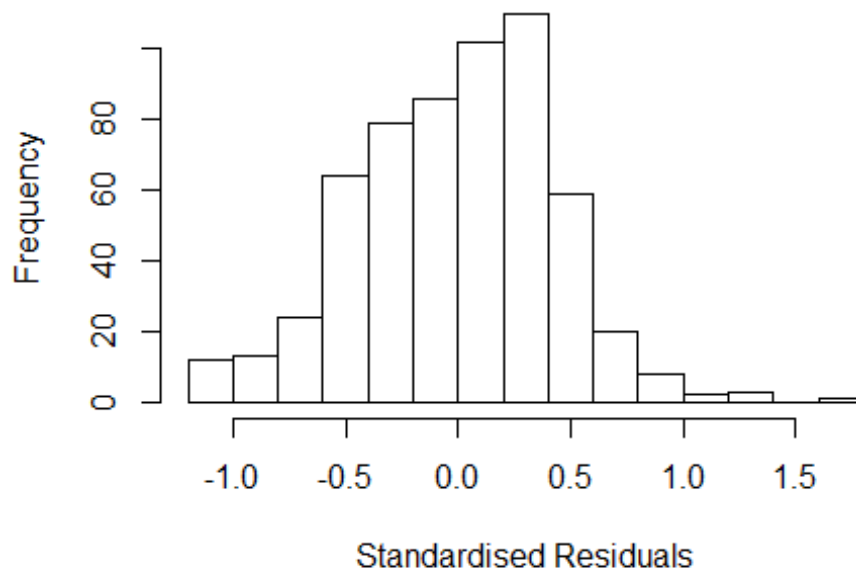
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1315 -0.2883  0.0308  0.2918  1.6823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97719    0.07959   12.28  <2e-16 ***
## FirstAuthorFemale1 0.01110    0.03811    0.29   0.771
## LastAuthorFemale1 0.00902    0.04910    0.18   0.854
## Year1997         0.14321    0.10304    1.39   0.165
## Year1998         0.06755    0.12122    0.56   0.578
## Year1999        -0.04018    0.09560   -0.42   0.674
## Year2000         0.07846    0.10452    0.75   0.453
## Year2001        -0.01867    0.11738   -0.16   0.874
## Year2002        -0.01031    0.10667   -0.10   0.923
## Year2003        -0.21135    0.11588   -1.82   0.069 .
## Year2004        -0.10725    0.11930   -0.90   0.369
## Year2005         0.08778    0.09621    0.91   0.362
```

```

## Year2006          0.03752    0.10353    0.36    0.717
## Year2007          0.08736    0.10407    0.84    0.402
## Year2008          0.12526    0.10131    1.24    0.217
## Year2009          0.06227    0.10718    0.58    0.561
## Year2010         -0.09995    0.13109   -0.76    0.446
## Year2011          0.01902    0.09990    0.19    0.849
## Year2012         -0.09560    0.11593   -0.82    0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0435, Adjusted R-squared:  0.013
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 540 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.108  0.887  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      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 1.088 1      1.043
## Year              1.088 16      1.003

```

Residuals from first author



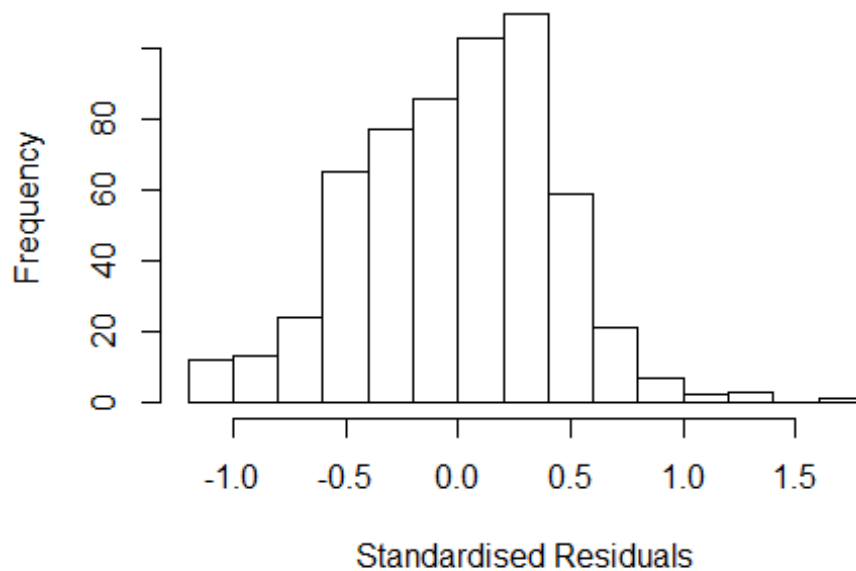
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1327 -0.2902  0.0294  0.2912  1.6803
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97836    0.07906   12.38  <2e-16 ***
## FirstAuthorFemale1 0.01170    0.03788    0.31   0.758
## Year1997        0.14269    0.10265    1.39   0.165
## Year1998        0.06822    0.12124    0.56   0.574
## Year1999       -0.04052    0.09550   -0.42   0.672
## Year2000        0.07760    0.10415    0.75   0.457
## Year2001       -0.02000    0.11693   -0.17   0.864
## Year2002       -0.00929    0.10675   -0.09   0.931
## Year2003       -0.21157    0.11591   -1.83   0.068 .
## Year2004       -0.10747    0.11940   -0.90   0.368
## Year2005        0.08800    0.09627    0.91   0.361
## Year2006        0.03831    0.10368    0.37   0.712
```

```

## Year2007          0.08781    0.10416    0.84    0.400
## Year2008          0.12535    0.10135    1.24    0.217
## Year2009          0.06211    0.10710    0.58    0.562
## Year2010         -0.09954    0.13092   -0.76    0.447
## Year2011          0.01971    0.10016    0.20    0.844
## Year2012         -0.09471    0.11585   -0.82    0.414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0435, Adjusted R-squared:  0.0147
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 537 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.107  0.887  0.952  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      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 1.068 1      1.033
## Year              1.068 16      1.002

```

Residuals from last author



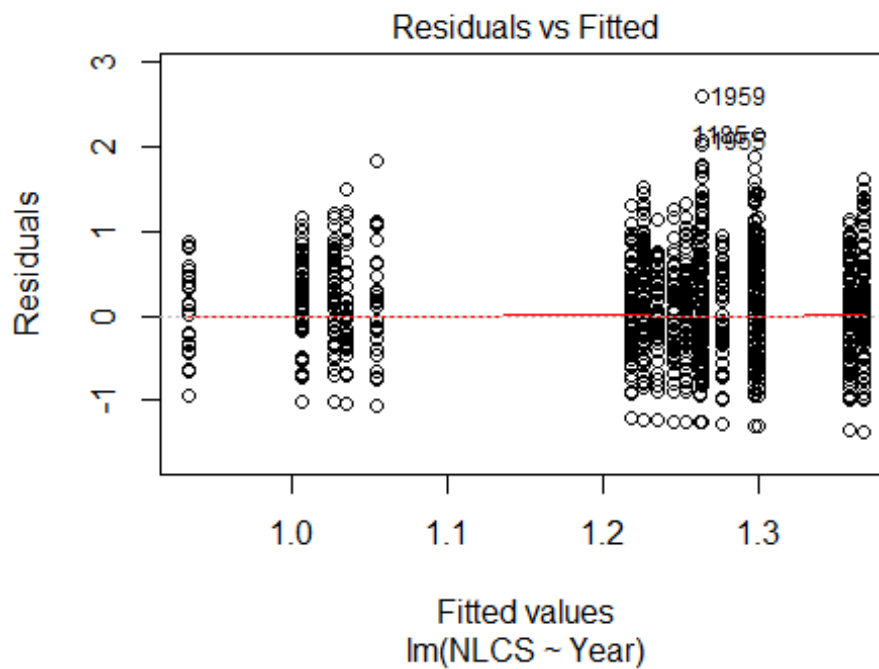
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1250 -0.2889 0.0282 0.2904 1.6809
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97985 0.07902 12.40 <2e-16 ***
## LastAuthorFemale1 0.01015 0.04882 0.21 0.835
## Year1997 0.14513 0.10262 1.41 0.158
## Year1998 0.06902 0.12107 0.57 0.569
## Year1999 -0.04061 0.09563 -0.42 0.671
## Year2000 0.07892 0.10435 0.76 0.450
## Year2001 -0.01715 0.11730 -0.15 0.884
## Year2002 -0.00948 0.10640 -0.09 0.929
## Year2003 -0.20990 0.11528 -1.82 0.069 .
## Year2004 -0.10658 0.11929 -0.89 0.372
## Year2005 0.09014 0.09548 0.94 0.345
## Year2006 0.03622 0.10334 0.35 0.726
```

```

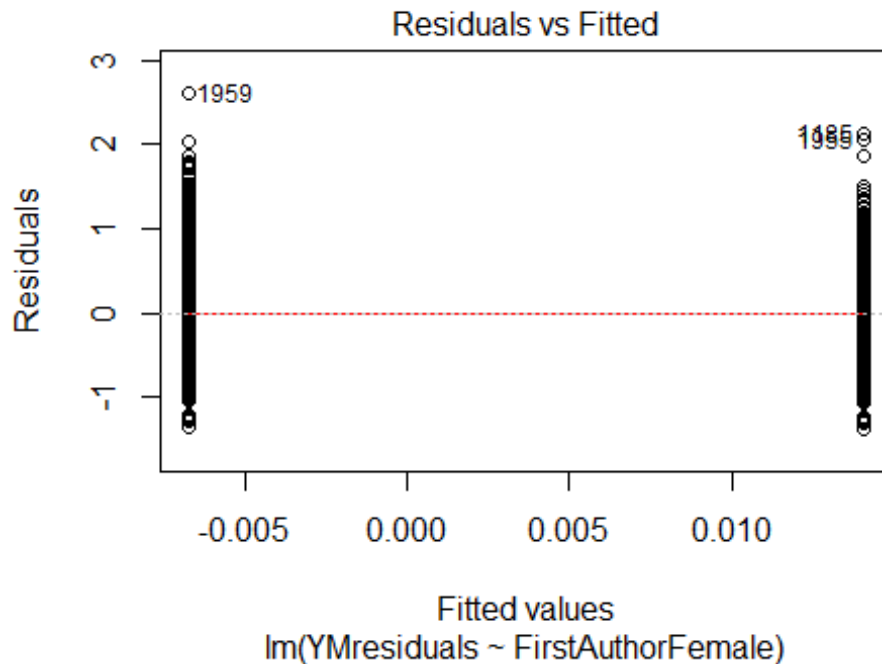
## Year2007      0.08771    0.10393    0.84    0.399
## Year2008      0.12583    0.10113    1.24    0.214
## Year2009      0.06299    0.10688    0.59    0.556
## Year2010     -0.10011    0.13113   -0.76    0.446
## Year2011      0.02218    0.09918    0.22    0.823
## Year2012     -0.09473    0.11578   -0.82    0.414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0434, Adjusted R-squared:  0.0146
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 541 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.107  0.887  0.952  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      1.72e-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 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
##   49   39   51   55   59   59   75   70   89   88  123  138  161  142  166
## 2011 2012
##  172  180
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   30   39   47   47   45   62   51   71   73   99  106  123  108  133
## 2011 2012

```

```
## 135 130
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 29 35 44 45 41 59 47 68 67 97 94 110 96 121
## 2011 2012
## 123 110
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 68, df = 16, p-value = 2e-08
```

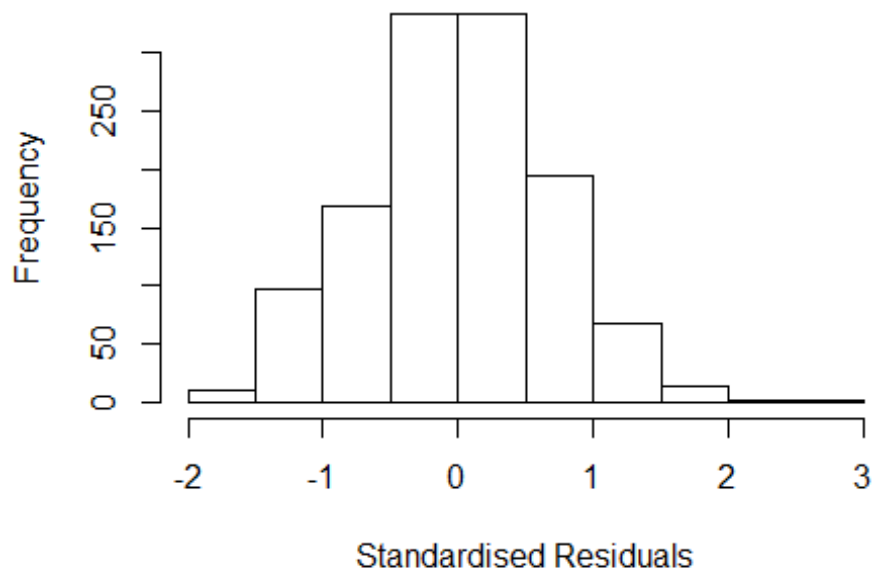


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.1, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.33753112841983"
## [1] "Male first author team size 2018 geometric mean: 1.60718170053732"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 4e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.20976203343073"
## [1] "Male last author team size 2018 geometric mean: 1.7151443586316"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, 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 1.644 1      1.282
## LastAuthorFemale  1.577 1      1.256
## UniqueAuthors     1.354 4      1.039
## Year              1.489 16      1.013
```


Residuals from first and last author and team size



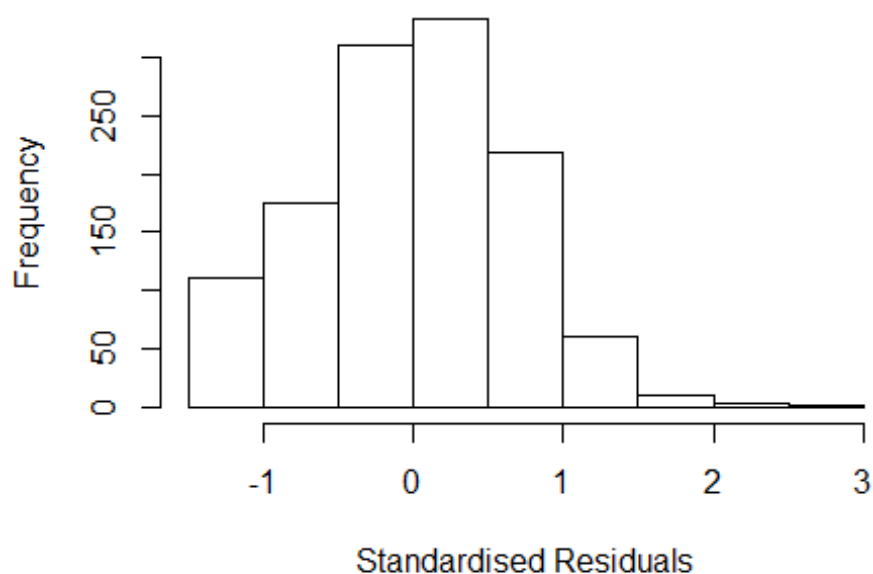
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1959 84864763916 3.861 2012      1400      1      2.645
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.673076 -0.432848  0.000432  0.466494  2.645319
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9305    0.1192    7.80 1.3e-14 ***
## FirstAuthorFemale1 -0.0219    0.0512   -0.43  0.6686
## LastAuthorFemale1  0.0175    0.0532    0.33  0.7421
## UniqueAuthors2     0.1941    0.0470    4.13 3.9e-05 ***
## UniqueAuthors3     0.4147    0.0573    7.24 7.9e-13 ***
## UniqueAuthors4     0.3515    0.1107    3.18  0.0015 **
## UniqueAuthors5     0.4383    0.2053    2.13  0.0330 *
## Year1997          -0.1190    0.1489   -0.80  0.4242
## Year1998          -0.0942    0.1525   -0.62  0.5368
## Year1999           0.1350    0.1427    0.95  0.3445
```

```

## Year2000          0.2582      0.1463      1.77      0.0778 .
## Year2001          0.2244      0.1597      1.41      0.1602
## Year2002         -0.0365      0.1442     -0.25      0.8004
## Year2003          0.1375      0.1398      0.98      0.3255
## Year2004          0.0152      0.1417      0.11      0.9149
## Year2005          0.1939      0.1437      1.35      0.1775
## Year2006          0.1478      0.1302      1.13      0.2567
## Year2007          0.2570      0.1363      1.89      0.0597 .
## Year2008          0.2241      0.1389      1.61      0.1069
## Year2009          0.2735      0.1414      1.93      0.0533 .
## Year2010          0.3043      0.1452      2.10      0.0364 *
## Year2011          0.1581      0.1404      1.13      0.2605
## Year2012          0.0911      0.1460      0.62      0.5329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.0795, Adjusted R-squared:  0.0626
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1093 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0719 0.8680 0.9460 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          8.18e-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 1.666 1      1.291
## LastAuthorFemale  1.632 1      1.278
## Year              1.137 16      1.004

```

Residuals from first and last author



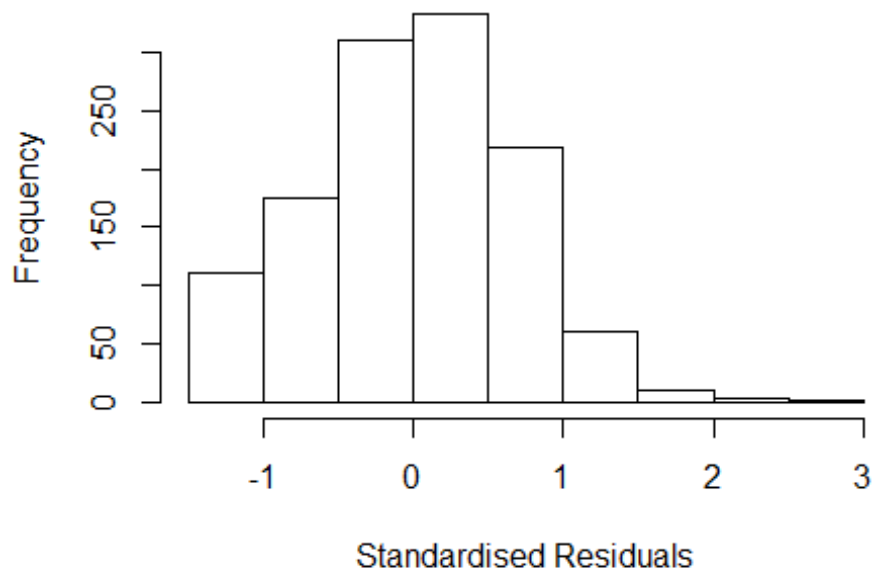
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1959 84864763916 3.861 2012    1400      1    2.697
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3739 -0.4459  0.0207  0.4797  2.6969
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.038594   0.123384   8.42  <2e-16 ***
## FirstAuthorFemale1  0.010441   0.053291   0.20   0.845
## LastAuthorFemale1  0.000259   0.056241   0.00   0.996
## Year1997        -0.131463   0.157261  -0.84   0.403
## Year1998        -0.051724   0.157948  -0.33   0.743
## Year1999         0.188197   0.147549   1.28   0.202
## Year2000         0.244508   0.149891   1.63   0.103
## Year2001         0.230435   0.162738   1.42   0.157
## Year2002         0.003018   0.149387   0.02   0.984
## Year2003         0.199103   0.145318   1.37   0.171
## Year2004         0.003710   0.147606   0.03   0.980
## Year2005         0.219958   0.147443   1.49   0.136
```

```

## Year2006          0.164722    0.134903    1.22    0.222
## Year2007          0.324606    0.141489    2.29    0.022 *
## Year2008          0.267431    0.143217    1.87    0.062 .
## Year2009          0.301967    0.146388    2.06    0.039 *
## Year2010          0.316521    0.149645    2.12    0.035 *
## Year2011          0.210988    0.145507    1.45    0.147
## Year2012          0.125556    0.151940    0.83    0.409
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.692
## Multiple R-squared:  0.029, Adjusted R-squared:  0.0145
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 1127 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0951 0.8710 0.9500 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      8.18e-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 1.085 1      1.042
## Year              1.085 16      1.003

```

Residuals from first author



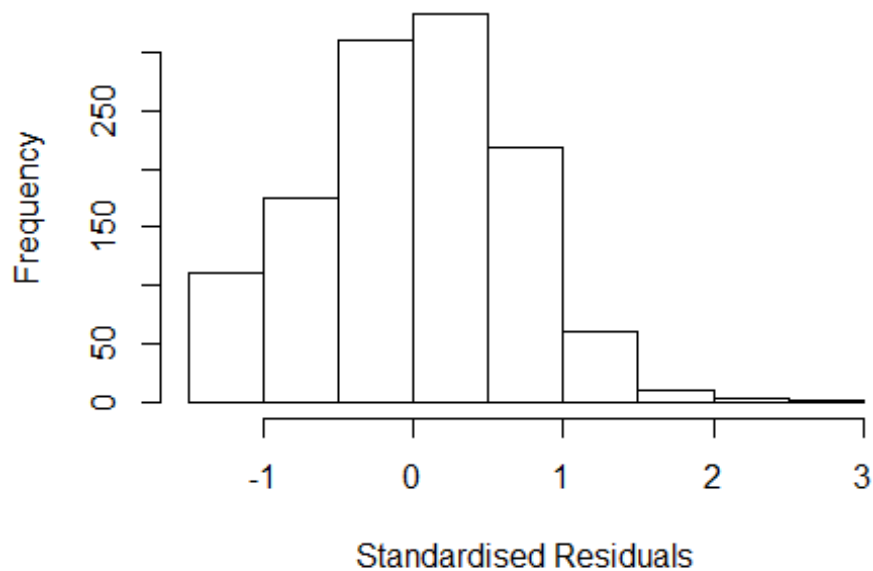
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1959 84864763916 3.861 2012    1400    1    2.697
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3738 -0.4459  0.0208  0.4796  2.6970
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03853    0.12330   8.42  <2e-16 ***
## FirstAuthorFemale1 0.01060    0.04301   0.25   0.805
## Year1997      -0.13138    0.15698  -0.84   0.403
## Year1998      -0.05168    0.15740  -0.33   0.743
## Year1999       0.18826    0.14715   1.28   0.201
## Year2000       0.24458    0.14983   1.63   0.103
## Year2001       0.23057    0.16249   1.42   0.156
## Year2002       0.00315    0.14900   0.02   0.983
## Year2003       0.19921    0.14484   1.38   0.169
## Year2004       0.00382    0.14681   0.03   0.979
## Year2005       0.22007    0.14655   1.50   0.133
## Year2006       0.16483    0.13429   1.23   0.220
```

```

## Year2007          0.32472    0.14113    2.30    0.022 *
## Year2008          0.26752    0.14258    1.88    0.061 .
## Year2009          0.30203    0.14613    2.07    0.039 *
## Year2010          0.31666    0.14946    2.12    0.034 *
## Year2011          0.21110    0.14530    1.45    0.147
## Year2012          0.12546    0.15063    0.83    0.405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.691
## Multiple R-squared:  0.029, Adjusted R-squared:  0.0153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 1127 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0941 0.8710 0.9500 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      8.18e-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 1.063 1          1.031
## Year            1.063 16          1.002

```

Residuals from last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1959 84864763916 3.861 2012    1400    1    2.697
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3718 -0.4474  0.0227  0.4780  2.6955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04034    0.12267   8.48  <2e-16 ***
## LastAuthorFemale1 0.00653    0.04536   0.14   0.886
## Year1997      -0.13268    0.15685  -0.85   0.398
## Year1998      -0.05226    0.15779  -0.33   0.741
## Year1999       0.18739    0.14724   1.27   0.203
## Year2000       0.24611    0.14984   1.64   0.101
## Year2001       0.22976    0.16280   1.41   0.158
## Year2002       0.00235    0.14914   0.02   0.987
## Year2003       0.19931    0.14520   1.37   0.170
## Year2004       0.00350    0.14746   0.02   0.981
## Year2005       0.22008    0.14739   1.49   0.136
## Year2006       0.16427    0.13468   1.22   0.223
```

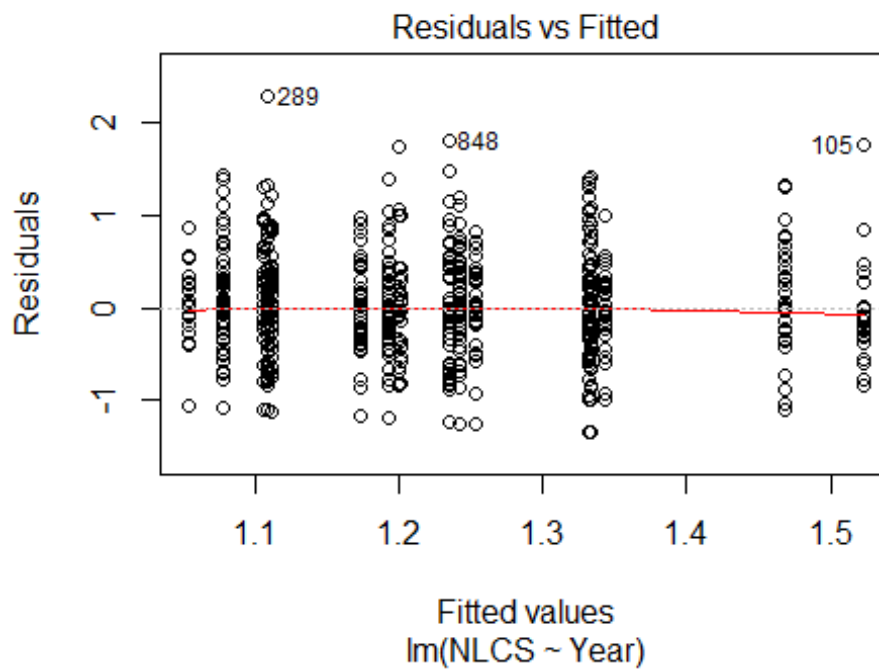
```

## Year2007      0.32497      0.14141      2.30      0.022 *
## Year2008      0.26768      0.14315      1.87      0.062 .
## Year2009      0.30251      0.14619      2.07      0.039 *
## Year2010      0.31614      0.14949      2.11      0.035 *
## Year2011      0.21097      0.14545      1.45      0.147
## Year2012      0.12517      0.15176      0.82      0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.691
## Multiple R-squared:  0.029, Adjusted R-squared:  0.0153
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0946 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      8.18e-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: 1223"
## [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
##   20   26   32   29   35   38   37   35   54   62   57   52   57   86   75
## 2011 2012
##   58   66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   22   24   24   30   33   32   29   43   49   49   46   49   67   63
## 2011 2012

```



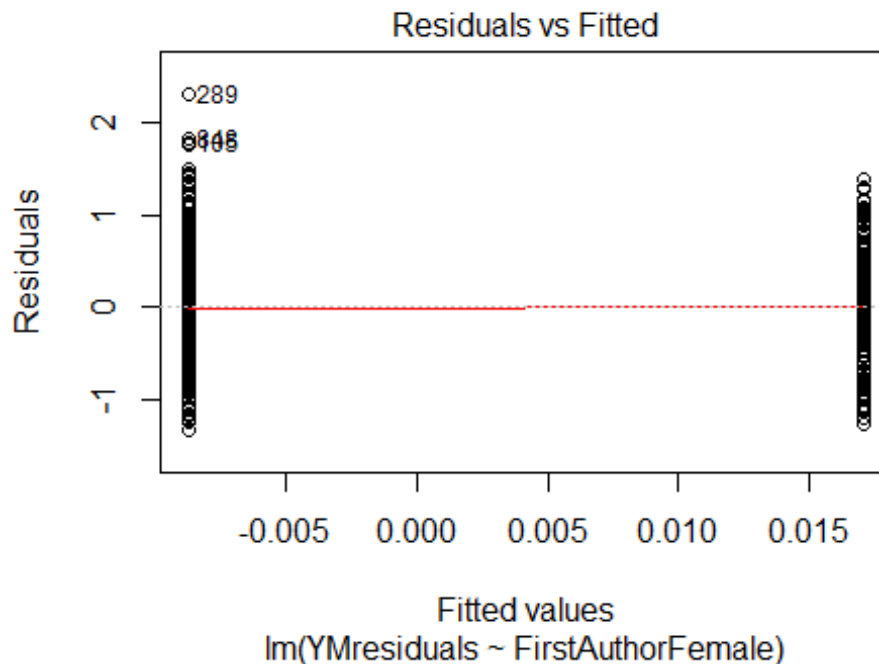
```
## 48 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 21 19 23 27 31 27 28 40 48 46 44 48 61 52
## 2011 2012
## 44 44
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.02
```



```
##
## 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: 2.64235367843914"
## [1] "Male first author team size 2018 geometric mean: 2.10754003396811"
## 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 = 450, p-value = 0.2
## 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: 2.25055868939235"

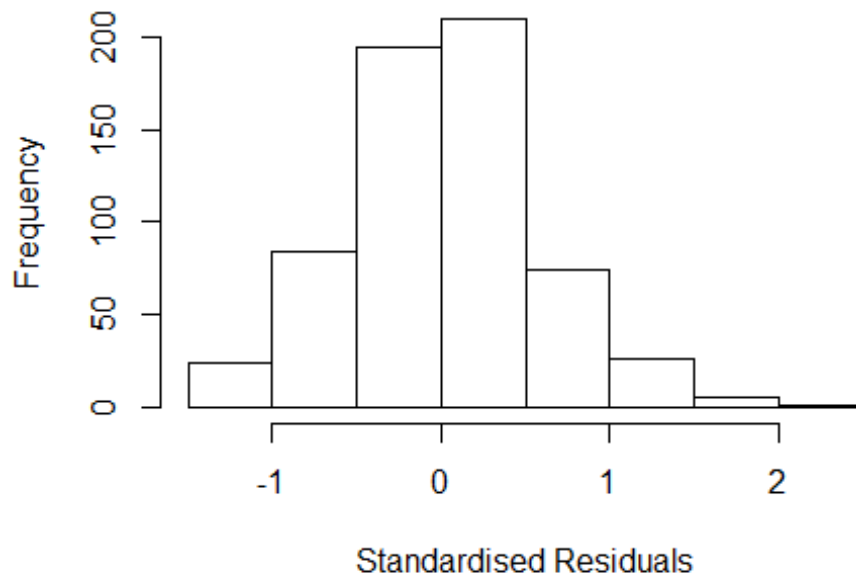
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 360, p-value = 0.7
## 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	1.607	1	1.268
LastAuthorFemale	1.450	1	1.204
UniqueAuthors	1.915	4	1.085
Year	2.297	16	1.026

Residuals from first and last author and team size



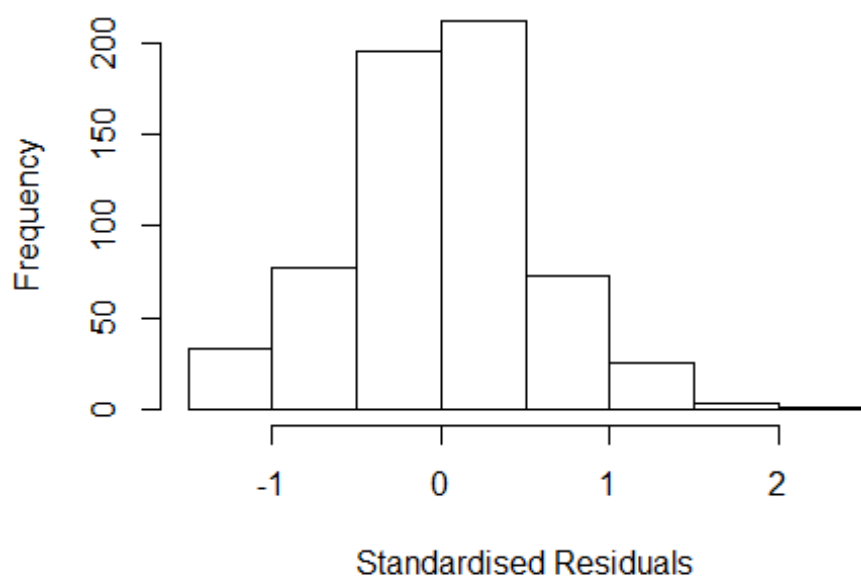
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40273 -0.35066 0.00636 0.33053 2.33423
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14105 0.12192 9.36 <2e-16 ***
## FirstAuthorFemale1 -0.02139 0.05694 -0.38 0.707
## LastAuthorFemale1 0.04870 0.05833 0.84 0.404
## UniqueAuthors2 0.16361 0.05275 3.10 0.002 **
## UniqueAuthors3 0.09259 0.07308 1.27 0.206
## UniqueAuthors4 0.23042 0.14341 1.61 0.109
## UniqueAuthors5 0.20267 0.20718 0.98 0.328
## Year1997 -0.10414 0.16157 -0.64 0.519
## Year1998 -0.24114 0.14932 -1.61 0.107
## Year1999 0.25684 0.16399 1.57 0.118
```

```

## Year2000          0.07077    0.14768    0.48    0.632
## Year2001          0.16230    0.17312    0.94    0.349
## Year2002          0.10953    0.14272    0.77    0.443
## Year2003         -0.05840    0.19016   -0.31    0.759
## Year2004         -0.23789    0.15875   -1.50    0.135
## Year2005          0.09807    0.16395    0.60    0.550
## Year2006          0.09249    0.14842    0.62    0.533
## Year2007         -0.04125    0.13839   -0.30    0.766
## Year2008         -0.03287    0.14504   -0.23    0.821
## Year2009         -0.16288    0.14358   -1.13    0.257
## Year2010          0.04806    0.13954    0.34    0.731
## Year2011         -0.00253    0.16942   -0.01    0.988
## Year2012         -0.12341    0.14705   -0.84    0.402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.0671, Adjusted R-squared:  0.0326
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 559 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0162 0.8670 0.9540 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          1.62e-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 1.458 1 1.208
## LastAuthorFemale 1.443 1 1.201
## Year 1.403 16 1.011

```

Residuals from first and last author



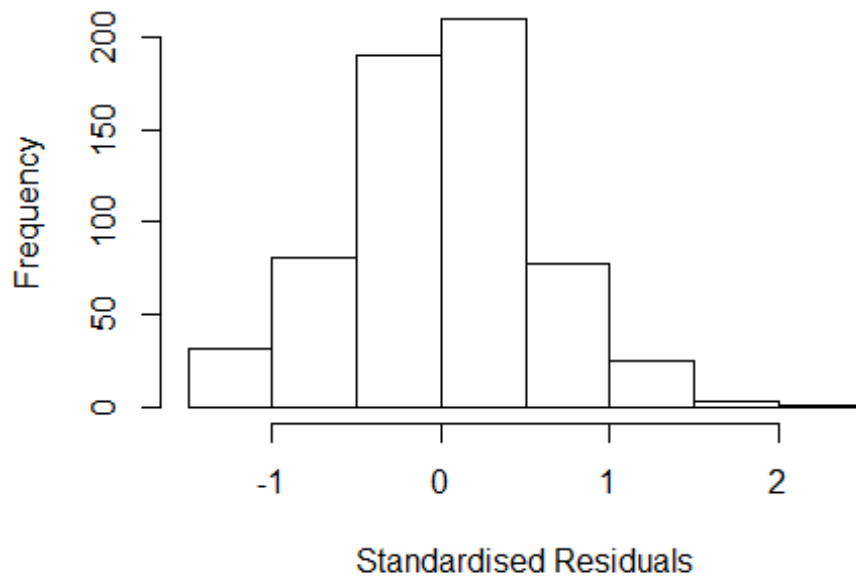
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32702 -0.33918 0.00848 0.34563 2.38955
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18041 0.11959 9.87 <2e-16 ***
## FirstAuthorFemale1 -0.00126 0.05544 -0.02 0.98
## LastAuthorFemale1 0.04456 0.05888 0.76 0.45
## Year1997 -0.11177 0.16095 -0.69 0.49
## Year1998 -0.22606 0.14646 -1.54 0.12
## Year1999 0.27310 0.15593 1.75 0.08 .
## Year2000 0.10111 0.14629 0.69 0.49
## Year2001 0.20956 0.16948 1.24 0.22
## Year2002 0.11385 0.13736 0.83 0.41
## Year2003 -0.01697 0.19282 -0.09 0.93
## Year2004 -0.16896 0.15470 -1.09 0.28
## Year2005 0.12981 0.16491 0.79 0.43
```

```

## Year2006          0.14661      0.14436      1.02      0.31
## Year2007          -0.01590      0.13800     -0.12      0.91
## Year2008          -0.01220      0.14092     -0.09      0.93
## Year2009          -0.12024      0.14115     -0.85      0.39
## Year2010           0.08037      0.13793      0.58      0.56
## Year2011           0.03702      0.16537      0.22      0.82
## Year2012          -0.05336      0.14054     -0.38      0.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0466, Adjusted R-squared:  0.018
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0066 0.8550 0.9540 0.8930 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      1.62e-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 1.148 1      1.071
## Year              1.148 16      1.004

```

Residuals from first author



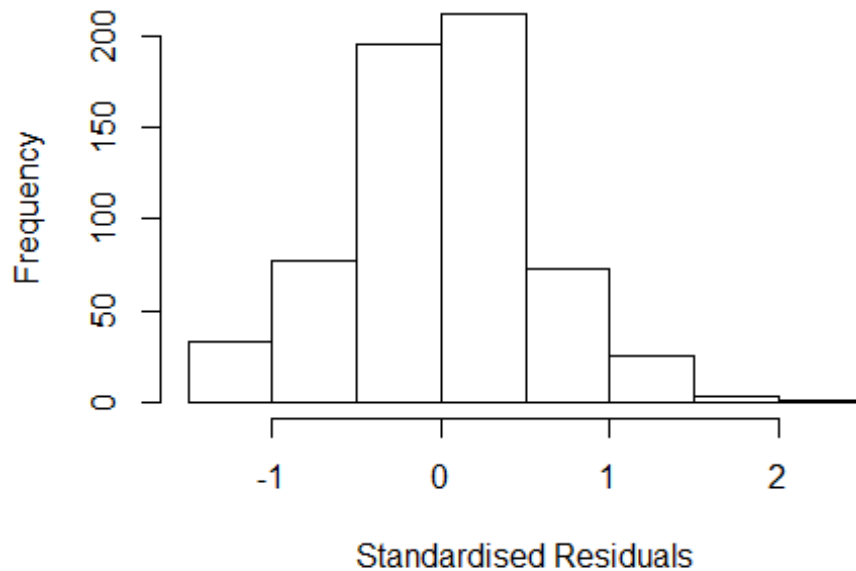
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33147 -0.33302  0.00561  0.34079  2.37360
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1887     0.1180   10.07  <2e-16 ***
## FirstAuthorFemale1  0.0199     0.0494    0.40   0.69
## Year1997        -0.1221     0.1597   -0.76   0.44
## Year1998        -0.2252     0.1449   -1.55   0.12
## Year1999         0.2702     0.1539    1.76   0.08 .
## Year2000         0.1006     0.1446    0.70   0.49
## Year2001         0.2032     0.1677    1.21   0.23
## Year2002         0.1157     0.1355    0.85   0.39
## Year2003        -0.0206     0.1906   -0.11   0.91
## Year2004        -0.1613     0.1521   -1.06   0.29
## Year2005         0.1360     0.1626    0.84   0.40
## Year2006         0.1428     0.1435    0.99   0.32
```

```

## Year2007          -0.0183      0.1362   -0.13      0.89
## Year2008          -0.0189      0.1394   -0.14      0.89
## Year2009          -0.1212      0.1397   -0.87      0.39
## Year2010           0.0798      0.1363     0.59      0.56
## Year2011           0.0328      0.1646     0.20      0.84
## Year2012          -0.0575      0.1395   -0.41      0.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0186
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 559 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0097 0.8540 0.9520 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      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 1.143 1          1.069
## Year            1.143 16          1.004

```


Residuals from last author



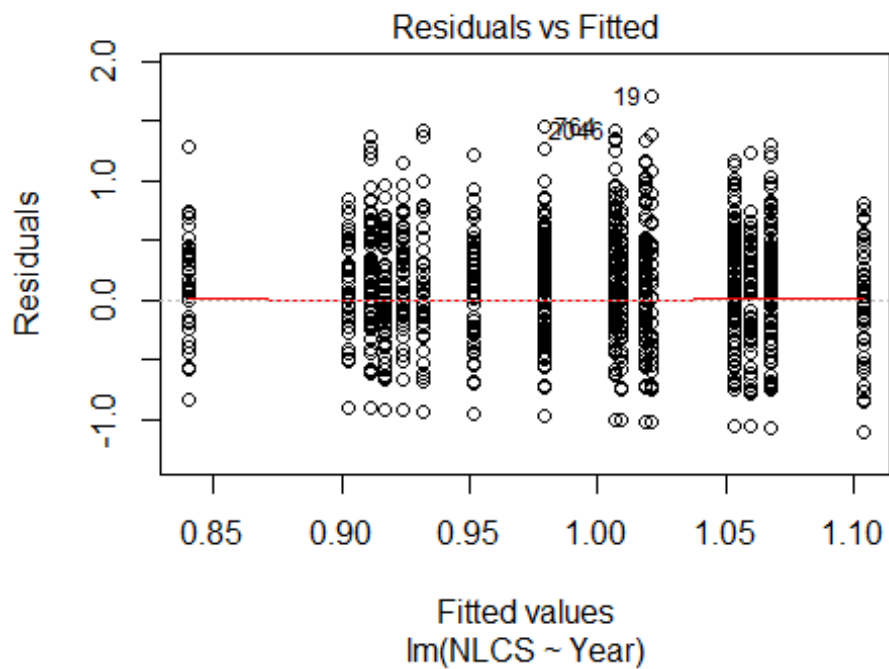
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3267 -0.3389 0.0086 0.3460 2.3898
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1802 0.1181 9.99 <2e-16 ***
## LastAuthorFemale1 0.0440 0.0525 0.84 0.40
## Year1997 -0.1121 0.1609 -0.70 0.49
## Year1998 -0.2260 0.1462 -1.55 0.12
## Year1999 0.2729 0.1559 1.75 0.08 .
## Year2000 0.1013 0.1461 0.69 0.49
## Year2001 0.2094 0.1696 1.23 0.22
## Year2002 0.1139 0.1371 0.83 0.41
## Year2003 -0.0169 0.1929 -0.09 0.93
## Year2004 -0.1690 0.1542 -1.10 0.27
## Year2005 0.1300 0.1629 0.80 0.43
## Year2006 0.1466 0.1444 1.01 0.31
```

```

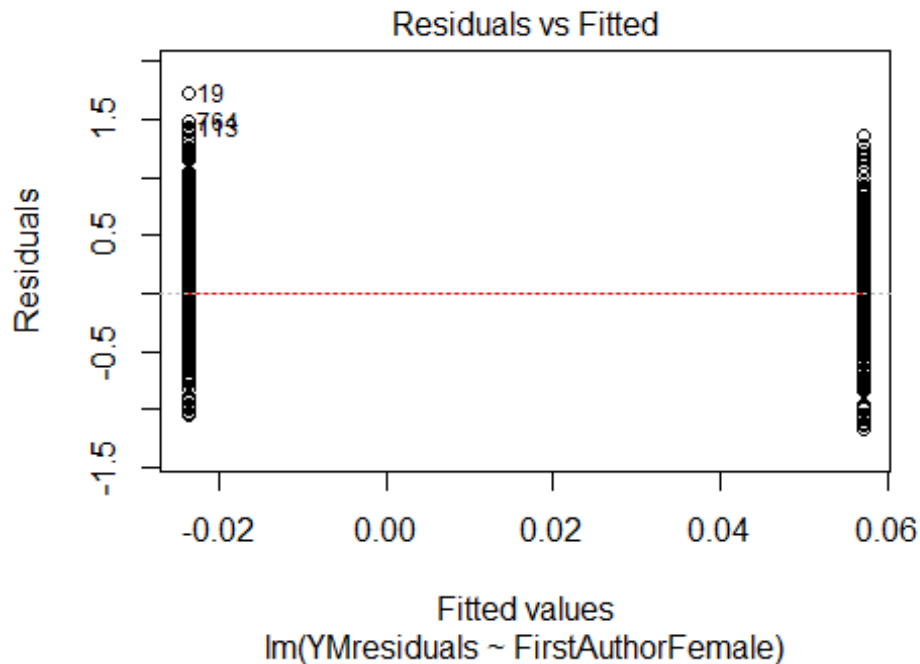
## Year2007          -0.0157      0.1374   -0.11      0.91
## Year2008          -0.0122      0.1411   -0.09      0.93
## Year2009          -0.1202      0.1409   -0.85      0.39
## Year2010           0.0805      0.1379    0.58      0.56
## Year2011           0.0368      0.1652    0.22      0.82
## Year2012          -0.0535      0.1406   -0.38      0.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.0467, Adjusted R-squared:  0.0197
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.006  0.855  0.953  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      1.62e-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: 619"
## [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
##   51   61   57   82   75   62   58   66   90   93  100   98  158  140  161
## 2011 2012
##  168  175
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   46   46   66   61   47   46   57   75   70   81   77  114  104  131
## 2011 2012

```

```
## 125 132
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 35 42 41 62 58 45 41 51 69 68 75 69 106 93 115
## 2011 2012
## 108 112
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```

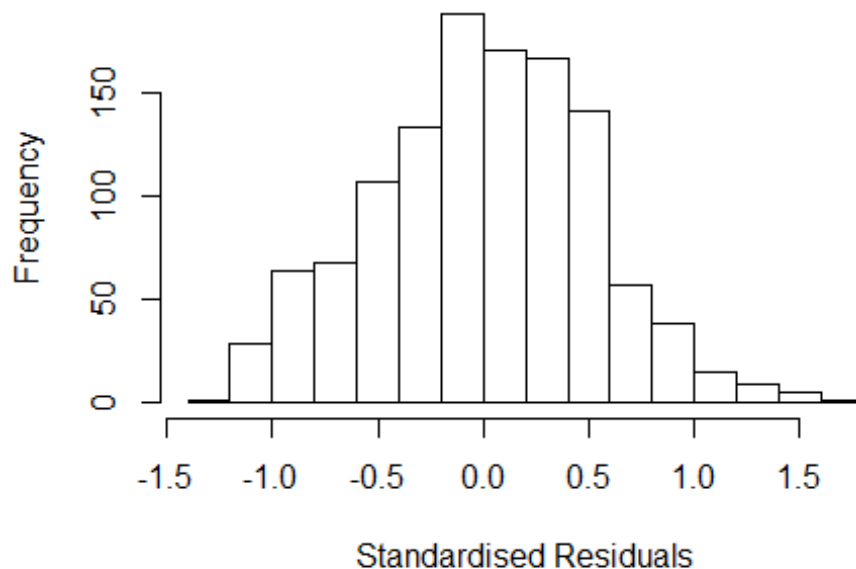


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.88, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.19853960250685"
## [1] "Male first author team size 2018 geometric mean: 1.83036673537463"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.07440022977065"
## [1] "Male last author team size 2018 geometric mean: 1.89775112395466"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 900, 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 1.381 1          1.175
## LastAuthorFemale  1.326 1          1.151
## UniqueAuthors     1.270 4          1.030
## Year               1.383 16         1.010
```

Residuals from first and last author and team size



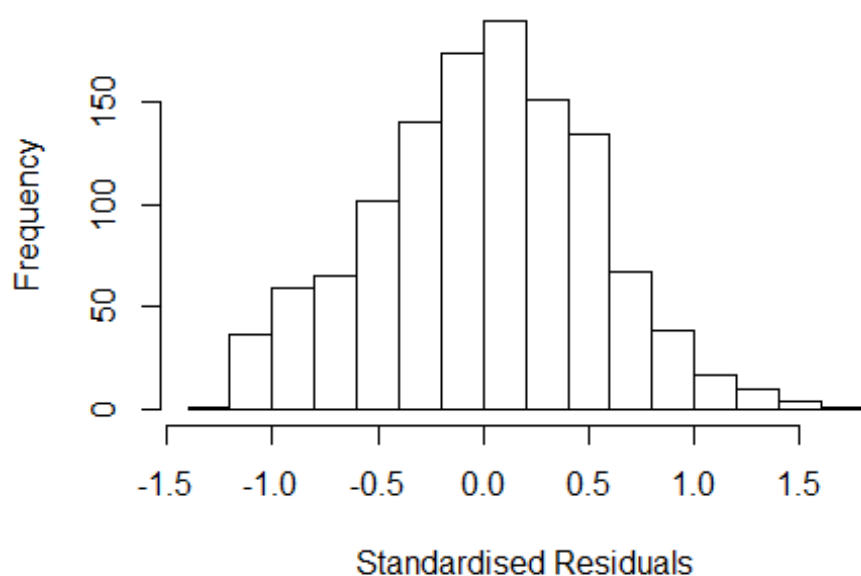
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20499 -0.35142  0.00776  0.35904  1.64473
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82854    0.12177   6.80 1.6e-11 ***
## FirstAuthorFemale1 0.02391    0.04059   0.59 0.55584
## LastAuthorFemale1 0.11125    0.04308   2.58 0.00993 **
## UniqueAuthors2    0.13748    0.03665   3.75 0.00018 ***
## UniqueAuthors3    0.21935    0.04415   4.97 7.7e-07 ***
## UniqueAuthors4    0.11400    0.09097   1.25 0.21037
## UniqueAuthors5   -0.05066    0.23601  -0.21 0.83007
## Year1997         -0.00624    0.14694  -0.04 0.96613
## Year1998          0.03030    0.13782   0.22 0.82601
## Year1999         -0.00071    0.13847  -0.01 0.99591
```

```

## Year2000      0.02807    0.14036    0.20  0.84154
## Year2001     -0.10809    0.14869   -0.73  0.46739
## Year2002      0.21053    0.14241    1.48  0.13959
## Year2003     -0.04643    0.13720   -0.34  0.73509
## Year2004     -0.03149    0.13844   -0.23  0.82009
## Year2005      0.01936    0.14168    0.14  0.89136
## Year2006      0.06697    0.13785    0.49  0.62719
## Year2007      0.07910    0.13387    0.59  0.55473
## Year2008      0.04833    0.13387    0.36  0.71816
## Year2009     -0.06735    0.13640   -0.49  0.62158
## Year2010      0.08553    0.13253    0.65  0.51882
## Year2011      0.10891    0.13522    0.81  0.42071
## Year2012      0.02194    0.13512    0.16  0.87106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.0554, Adjusted R-squared:  0.0376
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 101 weights are ~= 1. The remaining 1089 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.287  0.876  0.948  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      8.40e-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 1.337 1      1.156
## LastAuthorFemale  1.314 1      1.146
## Year              1.122 16      1.004

```

Residuals from first and last author



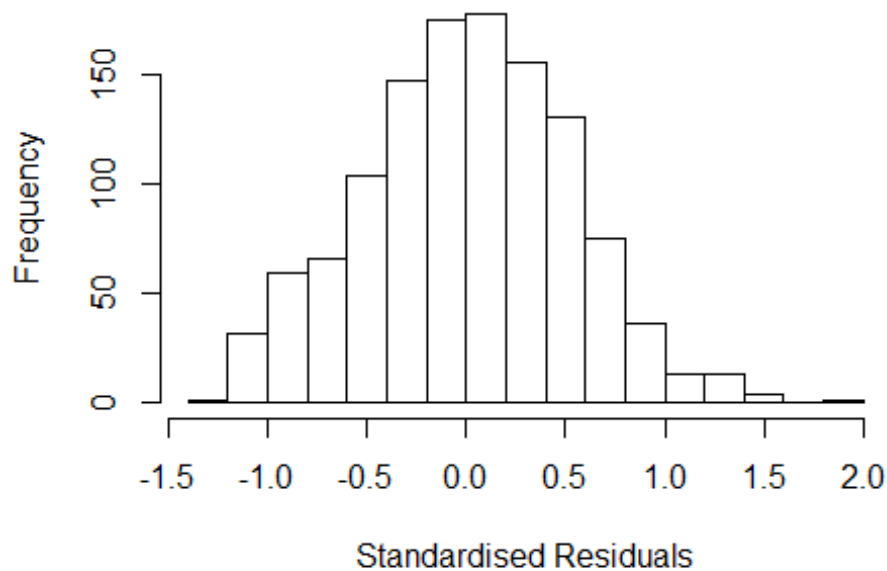
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.240 -0.347 0.016 0.361 1.723
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89588 0.12244 7.32 4.7e-13 ***
## FirstAuthorFemale1 0.03863 0.04045 0.96 0.340
## LastAuthorFemale1 0.10321 0.04338 2.38 0.018 *
## Year1997 -0.00607 0.14797 -0.04 0.967
## Year1998 0.03102 0.13699 0.23 0.821
## Year1999 -0.01067 0.13711 -0.08 0.938
## Year2000 0.03004 0.14080 0.21 0.831
## Year2001 -0.11341 0.14463 -0.78 0.433
## Year2002 0.20264 0.14390 1.41 0.159
## Year2003 -0.04298 0.13760 -0.31 0.755
## Year2004 -0.01390 0.13849 -0.10 0.920
## Year2005 0.04605 0.13978 0.33 0.742
```

```

## Year2006          0.10110    0.13832    0.73    0.465
## Year2007          0.09916    0.13466    0.74    0.462
## Year2008          0.06836    0.13246    0.52    0.606
## Year2009         -0.03671    0.13638   -0.27    0.788
## Year2010          0.12911    0.13209    0.98    0.329
## Year2011          0.13102    0.13481    0.97    0.331
## Year2012          0.04929    0.13579    0.36    0.717
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.015
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1082 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.872  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      8.40e-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 1.061 1      1.030
## Year              1.061 16      1.002

```


Residuals from first author



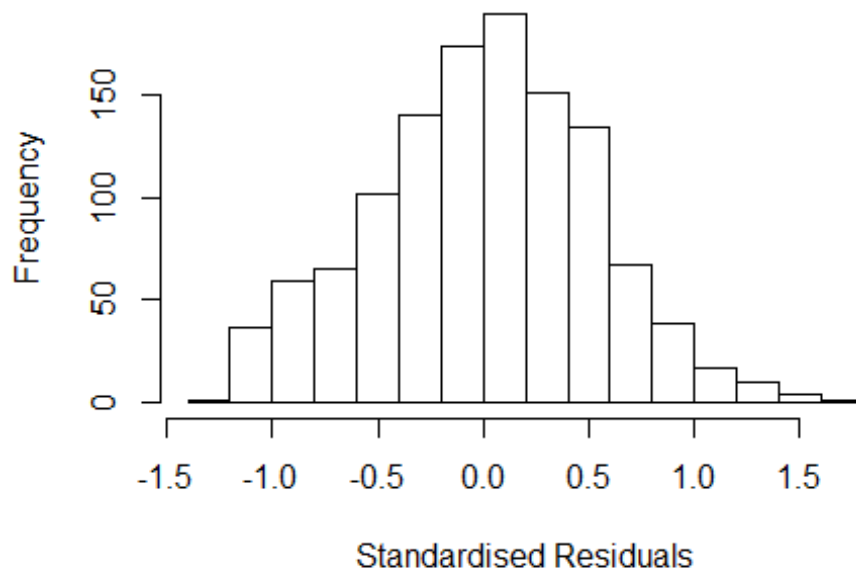
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2083 -0.3480 0.0131 0.3661 1.8210
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90100 0.12163 7.41 2.5e-13 ***
## FirstAuthorFemale1 0.08690 0.03611 2.41 0.016 *
## Year1997 0.00316 0.14800 0.02 0.983
## Year1998 0.04377 0.13725 0.32 0.750
## Year1999 -0.00958 0.13650 -0.07 0.944
## Year2000 0.03413 0.14002 0.24 0.807
## Year2001 -0.10894 0.14342 -0.76 0.448
## Year2002 0.22039 0.14345 1.54 0.125
## Year2003 -0.04466 0.13657 -0.33 0.744
## Year2004 -0.00402 0.13795 -0.03 0.977
## Year2005 0.05457 0.13918 0.39 0.695
## Year2006 0.09946 0.13785 0.72 0.471
```

```

## Year2007          0.10694    0.13421    0.80    0.426
## Year2008          0.08172    0.13189    0.62    0.536
## Year2009         -0.02872    0.13561   -0.21    0.832
## Year2010          0.13847    0.13158    1.05    0.293
## Year2011          0.14124    0.13379    1.06    0.291
## Year2012          0.05386    0.13519    0.40    0.690
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.0241, Adjusted R-squared:  0.00991
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1086 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.874  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      8.40e-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 1.044 1          1.022
## Year            1.044 16          1.001

```

Residuals from last author



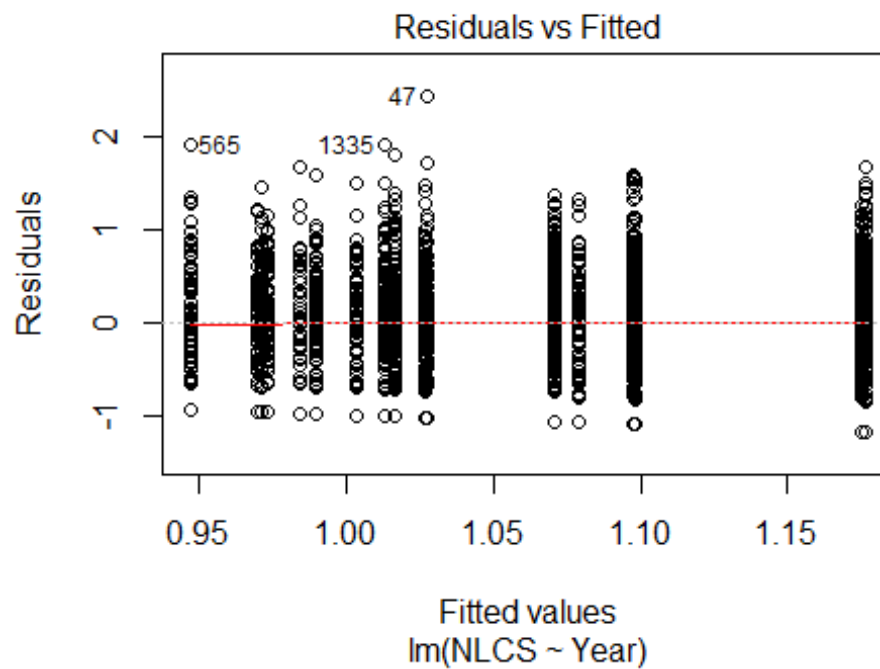
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2228 -0.3457 0.0139 0.3623 1.6982
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90124 0.12299 7.33 4.3e-13 ***
## LastAuthorFemale1 0.12255 0.03864 3.17 0.0016 **
## Year1997 -0.00821 0.14855 -0.06 0.9560
## Year1998 0.02756 0.13755 0.20 0.8412
## Year1999 -0.01315 0.13781 -0.10 0.9240
## Year2000 0.02805 0.14141 0.20 0.8428
## Year2001 -0.11484 0.14553 -0.79 0.4302
## Year2002 0.19905 0.14446 1.38 0.1685
## Year2003 -0.03870 0.13856 -0.28 0.7801
## Year2004 -0.01451 0.13912 -0.10 0.9169
## Year2005 0.04513 0.14048 0.32 0.7481
## Year2006 0.10383 0.13890 0.75 0.4549
```

```

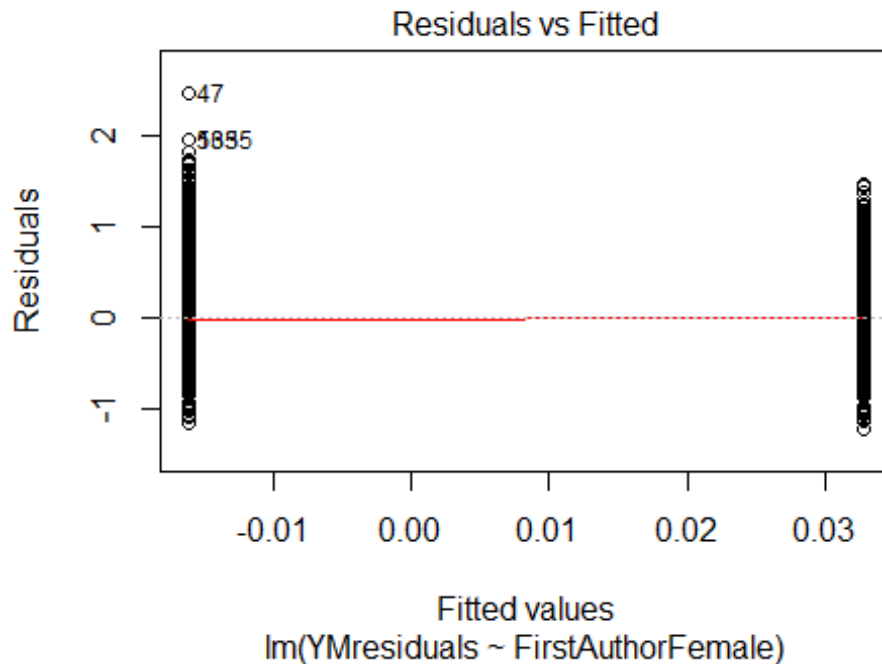
## Year2007      0.09959      0.13557      0.73      0.4627
## Year2008      0.07083      0.13320      0.53      0.5950
## Year2009     -0.03161      0.13706     -0.23      0.8176
## Year2010      0.13052      0.13284      0.98      0.3261
## Year2011      0.13076      0.13558      0.96      0.3350
## Year2012      0.05300      0.13648      0.39      0.6979
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.523
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.015
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1086 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.871  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      8.40e-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: 1190"
## [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
##   73   72   87   83  103  111  121  113  119  127  182  210  244  266  263
## 2011 2012
##  295  280
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   61   58   78   65   81   89  102   95   99  107  145  176  198  224  215
## 2011 2012

```

```
## 251 222
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 57 77 60 73 85 97 89 95 103 133 165 181 206 200
## 2011 2012
## 237 199
## [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
```

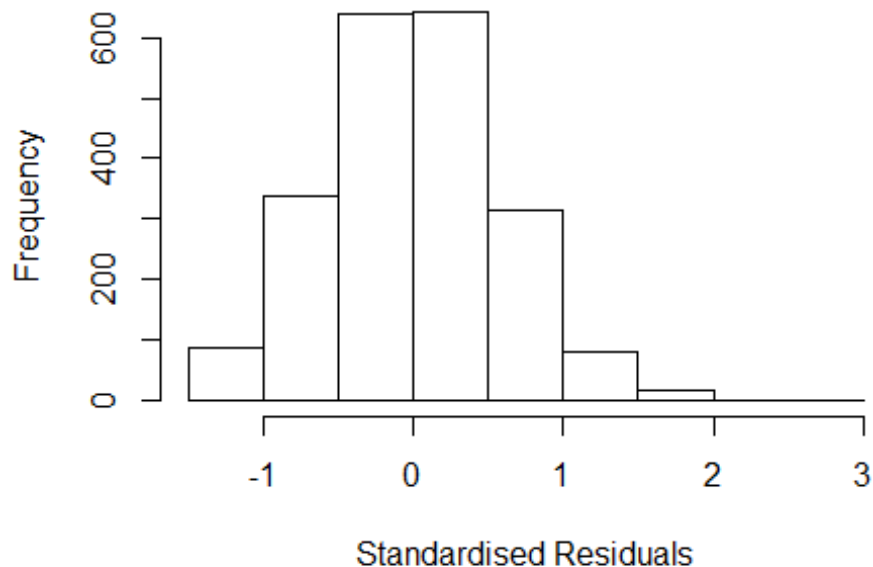


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.5, df = 1, p-value = 0.006
```



```
## [1] "Female first author team size 2018 geometric mean: 2.10580641895984"
## [1] "Male first author team size 2018 geometric mean: 1.91799649286432"
##
## 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] "Female last author team size 2018 geometric mean: 2.10464661664631"
## [1] "Male last author team size 2018 geometric mean: 1.91543619904137"
##
## 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 1.479 1          1.216
## LastAuthorFemale  1.445 1          1.202
## UniqueAuthors    1.206 4          1.024
## Year             1.304 16          1.008
```

Residuals from first and last author and team size



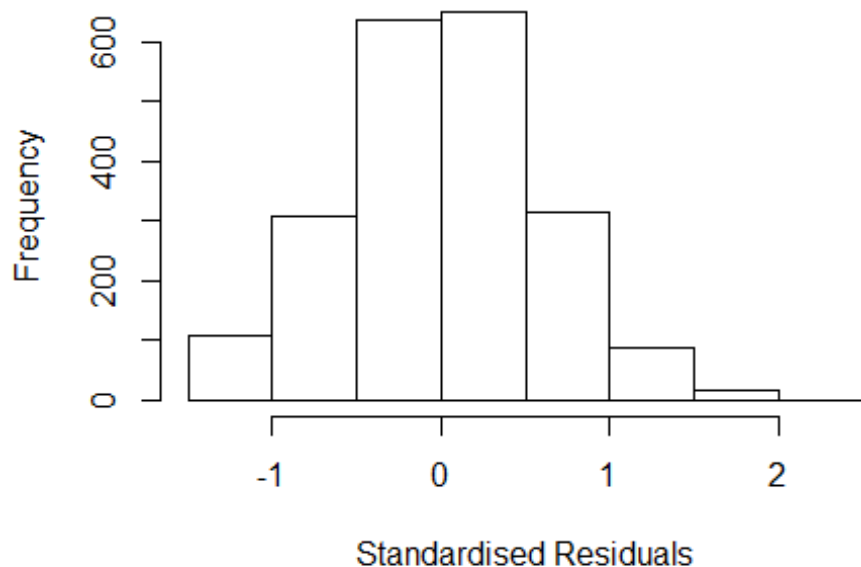
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 47 0030212023 3.466 1996    1403      2      2.54
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33546 -0.38438 -0.00455  0.39162  2.54033
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92567    0.08107   11.42 < 2e-16 ***
## FirstAuthorFemale1 0.04615    0.03252    1.42  0.15603
## LastAuthorFemale1 0.01443    0.03254    0.44  0.65746
## UniqueAuthors2    0.11405    0.03022    3.77  0.00017 ***
## UniqueAuthors3    0.17208    0.04104    4.19  2.9e-05 ***
## UniqueAuthors4    0.23166    0.08945    2.59  0.00967 **
## UniqueAuthors5    0.25147    0.11911    2.11  0.03488 *
## Year1997          0.00271    0.11216    0.02  0.98073
## Year1998          0.03004    0.10098    0.30  0.76610
## Year1999         -0.02014    0.10870   -0.19  0.85303
```

```

## Year2000      -0.06175      0.10306      -0.60      0.54911
## Year2001      -0.10959      0.10861      -1.01      0.31304
## Year2002       0.06558      0.10598       0.62      0.53613
## Year2003      -0.01999      0.09522      -0.21      0.83375
## Year2004       0.07346      0.09931       0.74      0.45954
## Year2005      -0.02154      0.09801      -0.22      0.82608
## Year2006      -0.03068      0.09637      -0.32      0.75027
## Year2007       0.01309      0.09078       0.14      0.88535
## Year2008       0.01171      0.09278       0.13      0.89959
## Year2009       0.05997      0.09101       0.66      0.51000
## Year2010       0.11917      0.09441       1.26      0.20698
## Year2011       0.03666      0.09116       0.40      0.68758
## Year2012       0.11756      0.09044       1.30      0.19381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0307, Adjusted R-squared:  0.0205
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 206 weights are ~= 1. The remaining 1911 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0224 0.8680 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          4.72e-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 1.480 1          1.217
## LastAuthorFemale  1.445 1          1.202
## Year              1.103 16          1.003

```


Residuals from first and last author



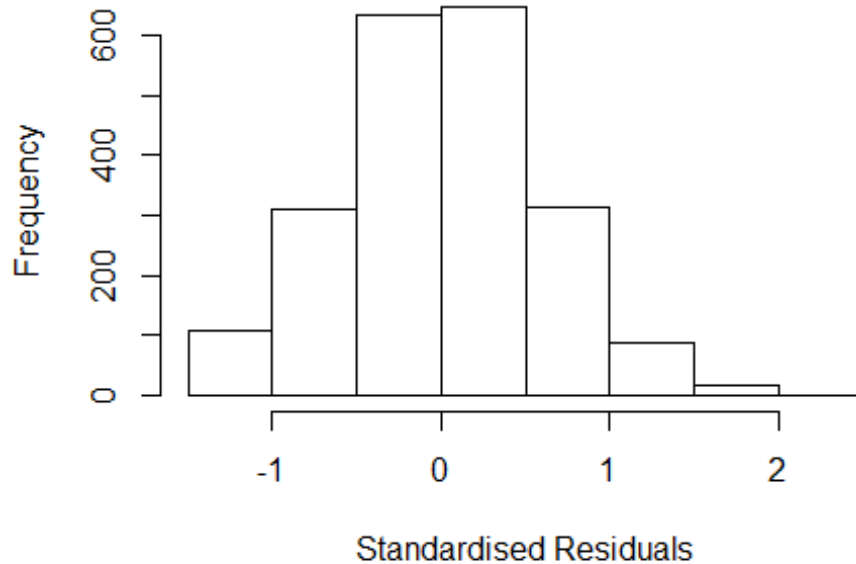
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19681 -0.39357 0.00338 0.39370 2.49741
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.968588 0.081184 11.93 <2e-16 ***
## FirstAuthorFemale1 0.049300 0.033009 1.49 0.135
## LastAuthorFemale1 0.013776 0.033045 0.42 0.677
## Year1997 -0.005789 0.112016 -0.05 0.959
## Year1998 0.016365 0.101136 0.16 0.871
## Year1999 -0.015542 0.109356 -0.14 0.887
## Year2000 -0.054581 0.103833 -0.53 0.599
## Year2001 -0.089756 0.108404 -0.83 0.408
## Year2002 0.086037 0.106072 0.81 0.417
## Year2003 -0.000700 0.096049 -0.01 0.994
## Year2004 0.086297 0.100139 0.86 0.389
## Year2005 0.007578 0.098314 0.08 0.939
```

```

## Year2006      0.000642    0.096741    0.01    0.995
## Year2007      0.045508    0.091191    0.50    0.618
## Year2008      0.027986    0.093568    0.30    0.765
## Year2009      0.086523    0.092383    0.94    0.349
## Year2010      0.165145    0.094225    1.75    0.080 .
## Year2011      0.078938    0.091529    0.86    0.389
## Year2012      0.161256    0.090646    1.78    0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.0158, Adjusted R-squared:  0.00739
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 1938 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0355 0.8670 0.9510 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      4.72e-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 1.076 1      1.037
## Year      1.076 16      1.002

```

Residuals from first author



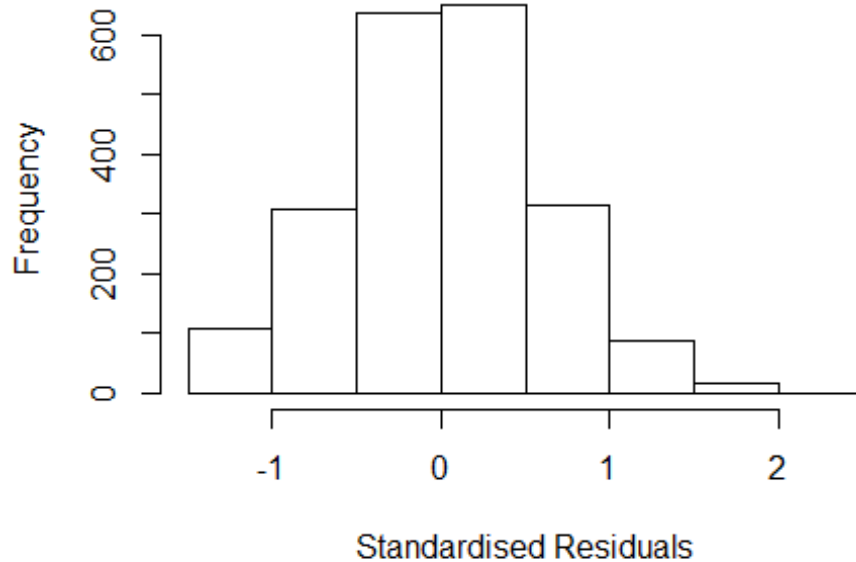
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19243 -0.39528 0.00408 0.39768 2.49562
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.70e-01 8.10e-02 11.98 <2e-16 ***
## FirstAuthorFemale1 5.68e-02 2.81e-02 2.02 0.044 *
## Year1997 -7.05e-03 1.12e-01 -0.06 0.950
## Year1998 1.67e-02 1.01e-01 0.17 0.869
## Year1999 -1.63e-02 1.09e-01 -0.15 0.882
## Year2000 -5.44e-02 1.04e-01 -0.52 0.601
## Year2001 -8.96e-02 1.08e-01 -0.83 0.409
## Year2002 8.57e-02 1.06e-01 0.81 0.419
## Year2003 -5.93e-05 9.61e-02 0.00 1.000
## Year2004 8.66e-02 1.00e-01 0.86 0.387
## Year2005 6.90e-03 9.83e-02 0.07 0.944
## Year2006 5.46e-04 9.67e-02 0.01 0.995
```

```

## Year2007          4.61e-02   9.12e-02   0.51   0.613
## Year2008          2.81e-02   9.35e-02   0.30   0.764
## Year2009          8.62e-02   9.24e-02   0.93   0.351
## Year2010          1.65e-01   9.42e-02   1.75   0.080 .
## Year2011          7.89e-02   9.15e-02   0.86   0.389
## Year2012          1.62e-01   9.06e-02   1.79   0.073 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.0157, Adjusted R-squared:  0.00778
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 1938 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0362 0.8670 0.9510 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      4.72e-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 1.051 1          1.025
## Year            1.051 16          1.002

```

Residuals from last author



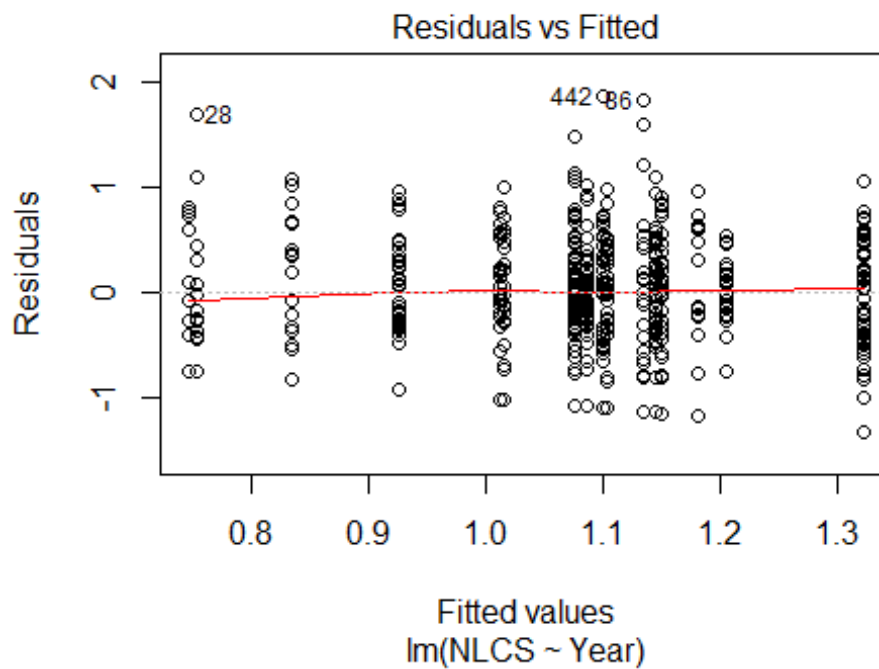
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18500 -0.39244 0.00473 0.39582 2.49367
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97233 0.08069 12.05 <2e-16 ***
## LastAuthorFemale1 0.04123 0.02813 1.47 0.143
## Year1997 -0.00424 0.11203 -0.04 0.970
## Year1998 0.01761 0.10058 0.18 0.861
## Year1999 -0.01550 0.10862 -0.14 0.887
## Year2000 -0.05172 0.10347 -0.50 0.617
## Year2001 -0.08490 0.10777 -0.79 0.431
## Year2002 0.08848 0.10565 0.84 0.402
## Year2003 0.00268 0.09560 0.03 0.978
## Year2004 0.08816 0.09994 0.88 0.378
## Year2005 0.00944 0.09806 0.10 0.923
## Year2006 0.00638 0.09639 0.07 0.947
```

```

## Year2007      0.04658      0.09080      0.51      0.608
## Year2008      0.03288      0.09318      0.35      0.724
## Year2009      0.09310      0.09190      1.01      0.311
## Year2010      0.17143      0.09380      1.83      0.068 .
## Year2011      0.08593      0.09096      0.94      0.345
## Year2012      0.16633      0.09014      1.85      0.065 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.00683
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 1945 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0366 0.8680 0.9510 0.9090 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      4.72e-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: 2117"
## [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
##   27   41   40   46   35   31   31   32   39   30   50   52   66   59   69
## 2011 2012
##   48   54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   26   27   23   15   13   20   23   32   21   36   38   44   42   46
## 2011 2012

```

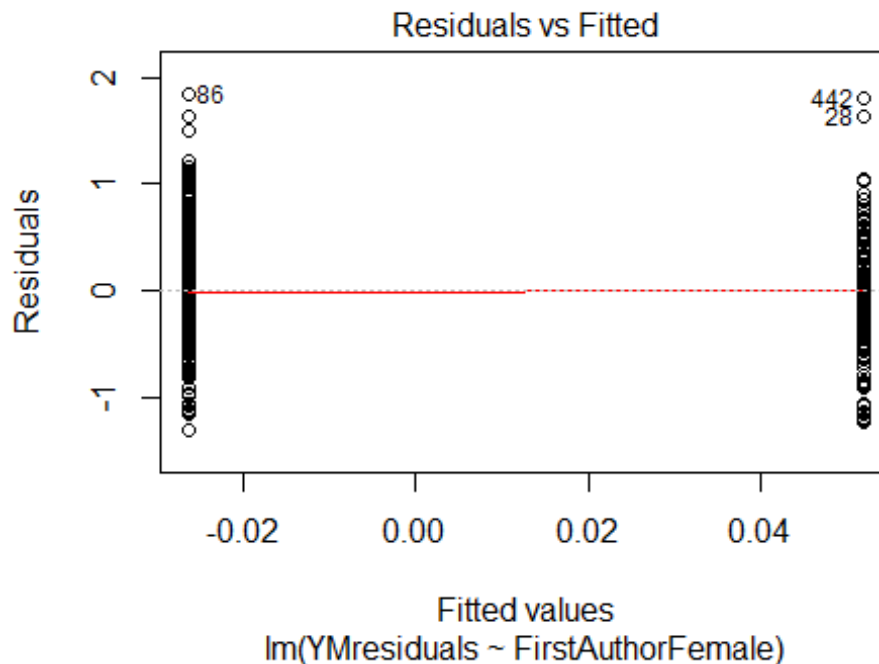
```
## 29 39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 26 23 20 12 11 18 19 27 20 28 35 39 40 42
## 2011 2012
## 25 30
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.023, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 1.8881750225898"
## [1] "Male first author team size 2018 geometric mean: 1.898132832128"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

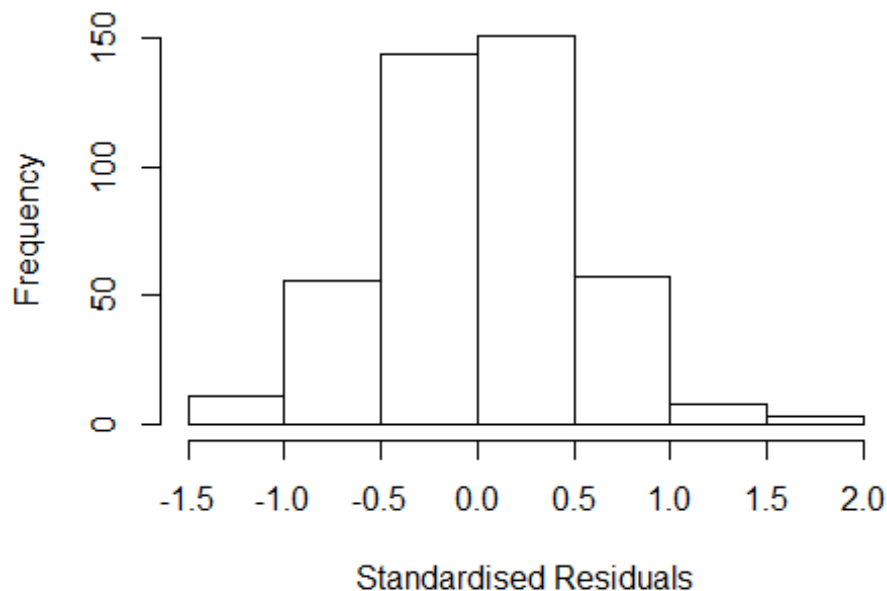
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 80, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.17619003622179"
## [1] "Male last author team size 2018 geometric mean: 1.78965424215283"

## 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 = 0.4
## 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 1.978 1      1.406
## LastAuthorFemale  1.779 1      1.334
## UniqueAuthors    2.386 4      1.115
## Year              2.979 16     1.035
```


Residuals from first and last author and team size



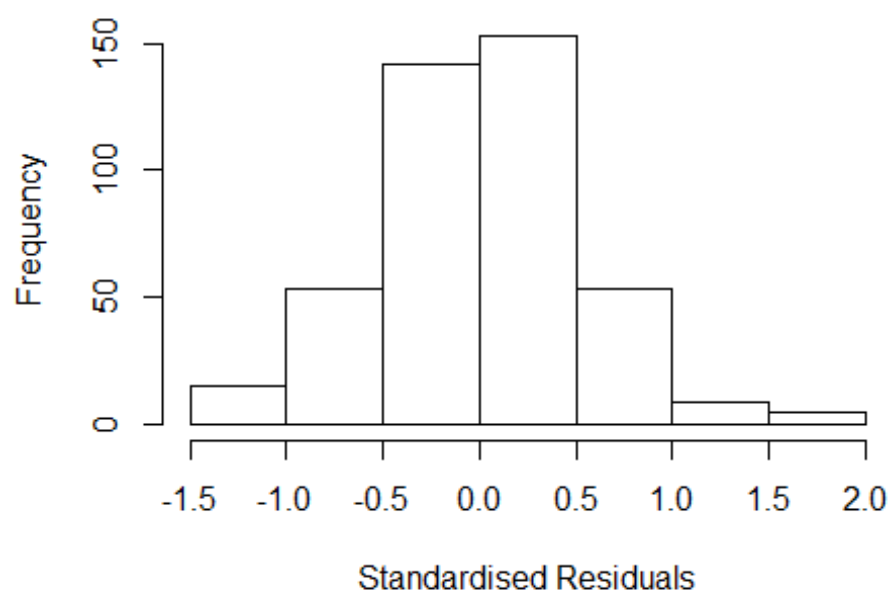
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23595 -0.32310 0.00517 0.34172 1.84504
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6315 0.1284 4.92 1.3e-06 ***
## FirstAuthorFemale1 0.1284 0.0684 1.88 0.06144 .
## LastAuthorFemale1 0.0259 0.0684 0.38 0.70522
## UniqueAuthors2 0.1372 0.0679 2.02 0.04397 *
## UniqueAuthors3 0.1258 0.0829 1.52 0.13007
## UniqueAuthors4 0.3969 0.1189 3.34 0.00092 ***
## UniqueAuthors5 0.6307 0.1840 3.43 0.00067 ***
## Year1997 0.3429 0.1917 1.79 0.07437 .
## Year1998 0.4695 0.2023 2.32 0.02083 *
## Year1999 0.0826 0.2014 0.41 0.68205
```

```

## Year2000          0.5462      0.2213      2.47  0.01401 *
## Year2001          0.0250      0.3204      0.08  0.93791
## Year2002          0.2836      0.1851      1.53  0.12634
## Year2003          0.2233      0.1911      1.17  0.24337
## Year2004          0.3478      0.1622      2.14  0.03260 *
## Year2005          0.3560      0.1546      2.30  0.02182 *
## Year2006          0.3865      0.1691      2.29  0.02279 *
## Year2007          0.1266      0.1634      0.77  0.43919
## Year2008          0.2763      0.1591      1.74  0.08310 .
## Year2009          0.3955      0.1503      2.63  0.00880 **
## Year2010          0.4657      0.1616      2.88  0.00417 **
## Year2011          0.3503      0.1823      1.92  0.05538 .
## Year2012          0.2081      0.1523      1.37  0.17242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.134, Adjusted R-squared:  0.0874
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 389 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.123  0.863  0.949   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      2.33e-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 1.884 1      1.373
## LastAuthorFemale  1.783 1      1.335
## Year              1.381 16      1.010

```

Residuals from first and last author



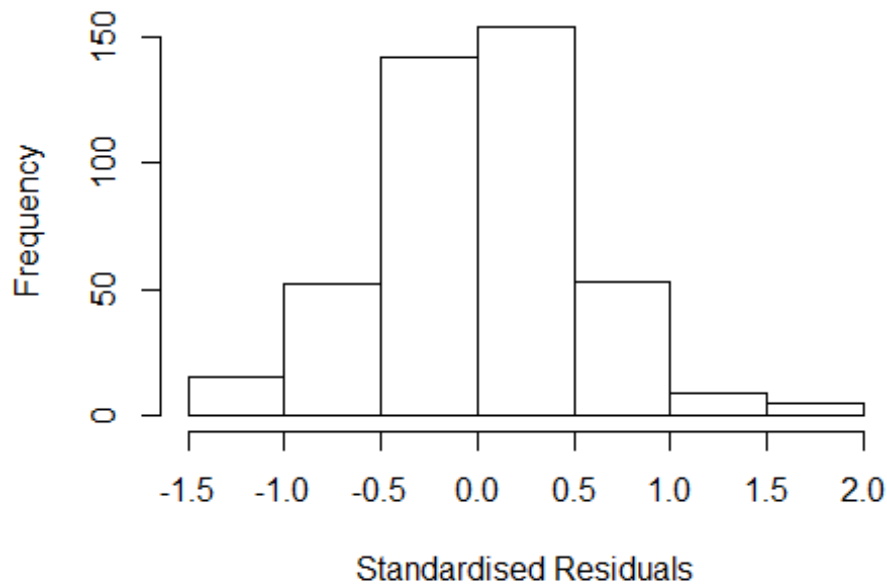
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2980 -0.3268 0.0119 0.3516 1.8155
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.66402 0.12488 5.32 1.7e-07 ***
## FirstAuthorFemale1 0.13578 0.06892 1.97 0.04950 *
## LastAuthorFemale1 -0.00749 0.07090 -0.11 0.91591
## Year1997 0.37369 0.18146 2.06 0.04009 *
## Year1998 0.46649 0.19244 2.42 0.01578 *
## Year1999 0.10777 0.20739 0.52 0.60361
## Year2000 0.63399 0.24290 2.61 0.00938 **
## Year2001 0.08112 0.30185 0.27 0.78826
## Year2002 0.32916 0.18359 1.79 0.07373 .
## Year2003 0.32203 0.18028 1.79 0.07480 .
## Year2004 0.43423 0.15389 2.82 0.00501 **
## Year2005 0.48802 0.14165 3.45 0.00063 ***
```

```

## Year2006          0.51387    0.16507    3.11  0.00198 **
## Year2007          0.22414    0.15288    1.47  0.14336
## Year2008          0.34068    0.15542    2.19  0.02894 *
## Year2009          0.47632    0.14417    3.30  0.00104 **
## Year2010          0.56387    0.15274    3.69  0.00025 ***
## Year2011          0.43974    0.18318    2.40  0.01681 *
## Year2012          0.29769    0.14830    2.01  0.04537 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0971, Adjusted R-squared:  0.0575
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 388 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.135  0.849   0.948   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      2.33e-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 1.226 1         1.107
## Year              1.226 16         1.006

```

Residuals from first author



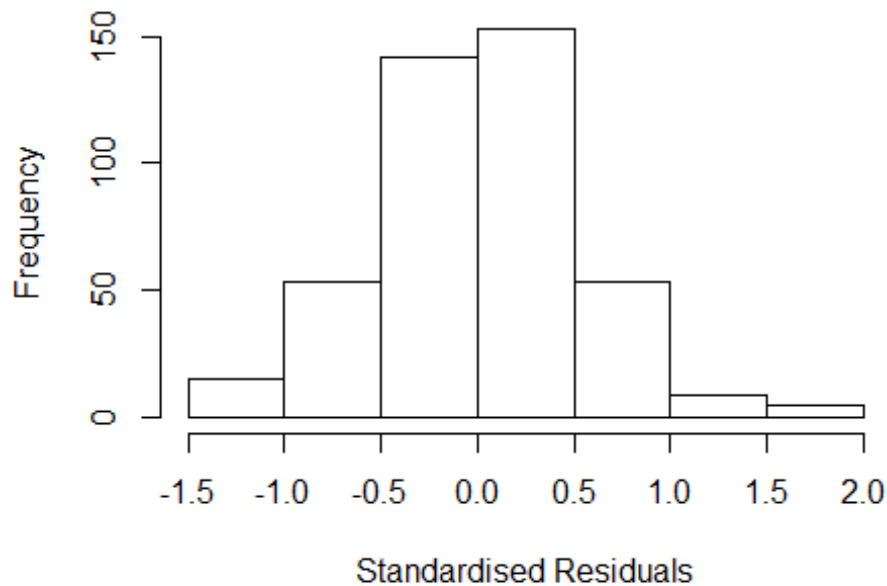
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.299 -0.326 0.011 0.352 1.817
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6633 0.1248 5.31 1.8e-07 ***
## FirstAuthorFemale1 0.1321 0.0560 2.36 0.01885 *
## Year1997 0.3746 0.1815 2.06 0.03964 *
## Year1998 0.4655 0.1920 2.42 0.01577 *
## Year1999 0.1074 0.2090 0.51 0.60766
## Year2000 0.6362 0.2431 2.62 0.00919 **
## Year2001 0.0819 0.3068 0.27 0.78970
## Year2002 0.3295 0.1839 1.79 0.07397 .
## Year2003 0.3220 0.1793 1.80 0.07328 .
## Year2004 0.4343 0.1535 2.83 0.00488 **
## Year2005 0.4870 0.1410 3.46 0.00061 ***
## Year2006 0.5137 0.1657 3.10 0.00206 **
```

```

## Year2007          0.2232      0.1525      1.46  0.14416
## Year2008          0.3399      0.1551      2.19  0.02899 *
## Year2009          0.4759      0.1434      3.32  0.00098 ***
## Year2010          0.5645      0.1533      3.68  0.00026 ***
## Year2011          0.4397      0.1839      2.39  0.01723 *
## Year2012          0.2973      0.1479      2.01  0.04516 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.0973, Adjusted R-squared:  0.0601
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 388 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.131  0.846  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      2.33e-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 1.148 1      1.072
## Year      1.148 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
## -1.346029 -0.321429 0.000697 0.351228 1.787903
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6795 0.1248 5.45 8.9e-08 ***
## LastAuthorFemale1 0.0656 0.0567 1.16 0.24831
## Year1997 0.3866 0.1807 2.14 0.03296 *
## Year1998 0.4786 0.1958 2.44 0.01496 *
## Year1999 0.1005 0.2027 0.50 0.62021
## Year2000 0.6665 0.2460 2.71 0.00701 **
## Year2001 0.1156 0.3070 0.38 0.70661
## Year2002 0.3709 0.1786 2.08 0.03846 *
## Year2003 0.3169 0.1768 1.79 0.07377 .
## Year2004 0.4617 0.1527 3.02 0.00266 **
## Year2005 0.4814 0.1428 3.37 0.00082 ***
## Year2006 0.5253 0.1639 3.20 0.00146 **
```

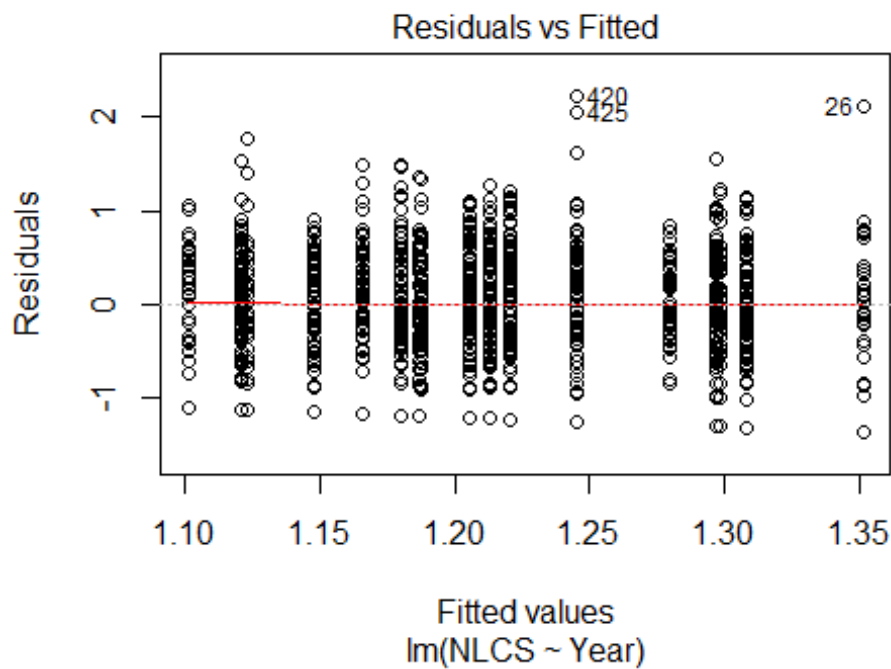
```

## Year2007          0.2124      0.1531      1.39  0.16595
## Year2008          0.3517      0.1537      2.29  0.02266 *
## Year2009          0.4832      0.1443      3.35  0.00089 ***
## Year2010          0.5937      0.1512      3.93  0.00010 ***
## Year2011          0.4659      0.1834      2.54  0.01143 *
## Year2012          0.2963      0.1466      2.02  0.04392 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0866, Adjusted R-squared:  0.0489
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 393 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.858  0.952  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      2.33e-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: 430"
## [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
##   41   54   59   52   86  104   70   68   57   94   88   98  110  148  112
## 2011 2012
##  124  142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   39   41   40   57   75   59   54   45   81   68   86   91  129   91
## 2011 2012

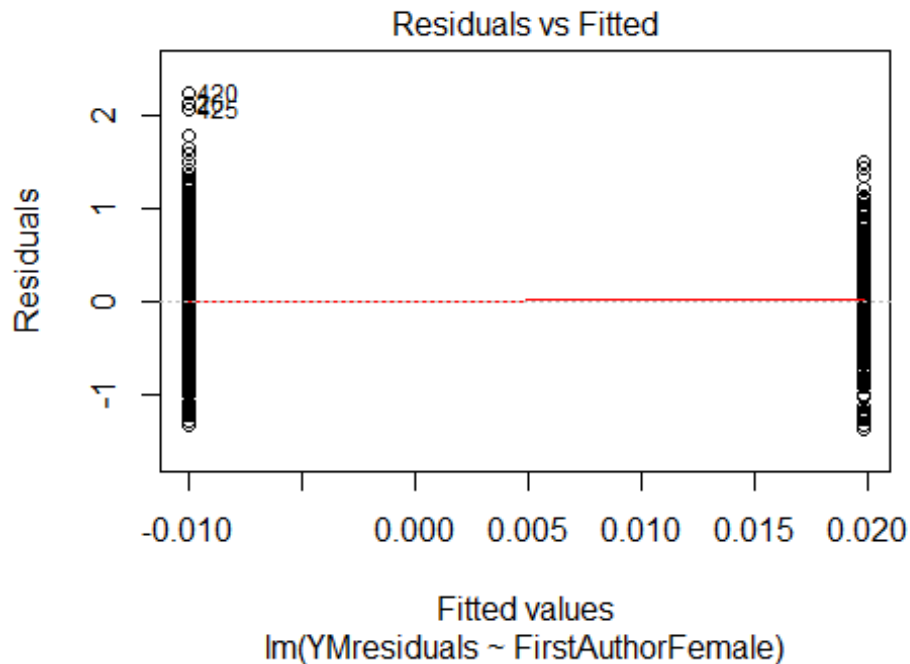
```



```
## 97 114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 38 37 38 55 71 52 47 43 76 61 75 85 117 87
## 2011 2012
## 91 94
## [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
```



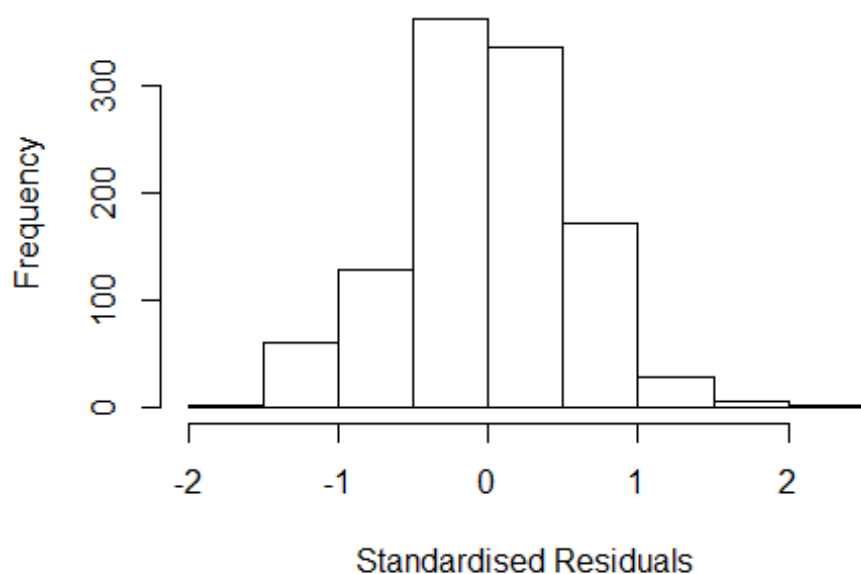
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 6e-05
## [1] "Female first author team size 2018 geometric mean: 2.09174160218546"
## [1] "Male first author team size 2018 geometric mean: 1.82459528412491"
## 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 = 920, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.95992436953563"
## [1] "Male last author team size 2018 geometric mean: 1.92061569724815"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 770, 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	1.403	1	1.185
LastAuthorFemale	1.403	1	1.184
UniqueAuthors	1.471	4	1.049
Year	1.503	16	1.013

Residuals from first and last author and team size



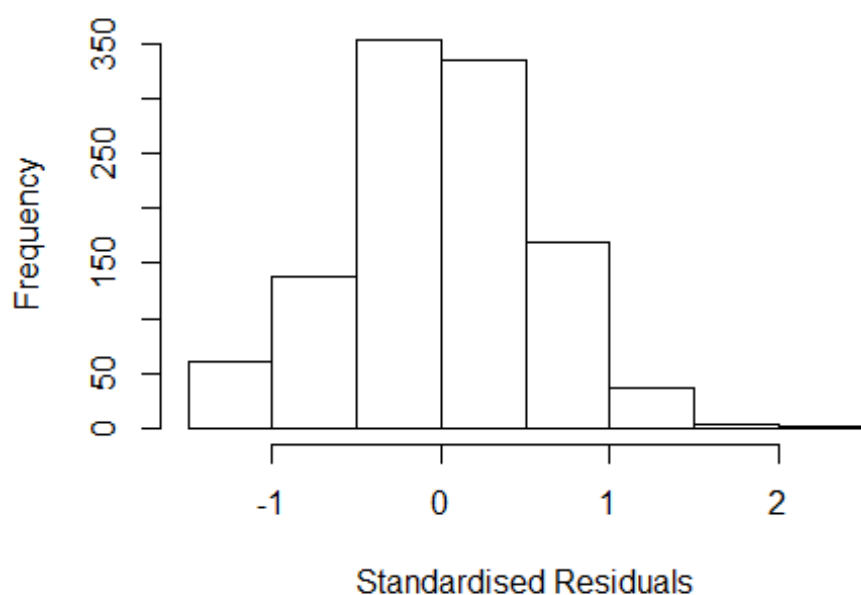
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.51365 -0.35703 -0.00517 0.36557 2.16923
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3087 0.1364 9.60 < 2e-16 ***
## FirstAuthorFemale1 0.0349 0.0396 0.88 0.37832
## LastAuthorFemale1 -0.0222 0.0415 -0.54 0.59199
## UniqueAuthors2 0.1412 0.0391 3.61 0.00032 ***
## UniqueAuthors3 0.2049 0.0554 3.70 0.00023 ***
## UniqueAuthors4 0.3104 0.0823 3.77 0.00017 ***
## UniqueAuthors5 -0.4131 0.1486 -2.78 0.00552 **
## Year1997 -0.2734 0.1703 -1.61 0.10862
## Year1998 -0.2436 0.1622 -1.50 0.13336
## Year1999 -0.3138 0.1774 -1.77 0.07719 .
```

```

## Year2000          -0.2034      0.1527    -1.33   0.18314
## Year2001          -0.1531      0.1628    -0.94   0.34708
## Year2002          -0.2237      0.1729    -1.29   0.19599
## Year2003          -0.1124      0.1490    -0.75   0.45072
## Year2004          -0.0984      0.1591    -0.62   0.53651
## Year2005          -0.1254      0.1473    -0.85   0.39467
## Year2006          -0.2344      0.1500    -1.56   0.11847
## Year2007          -0.0870      0.1508    -0.58   0.56397
## Year2008          -0.2012      0.1527    -1.32   0.18794
## Year2009          -0.2856      0.1456    -1.96   0.05004 .
## Year2010          -0.2601      0.1524    -1.71   0.08823 .
## Year2011          -0.1736      0.1540    -1.13   0.25975
## Year2012          -0.2345      0.1497    -1.57   0.11735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.05,   Adjusted R-squared:  0.0305
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 1006 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0847 0.8690 0.9530 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          9.12e-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 1.364 1          1.168
## LastAuthorFemale 1.402 1          1.184
## Year          1.102 16          1.003

```

Residuals from first and last author



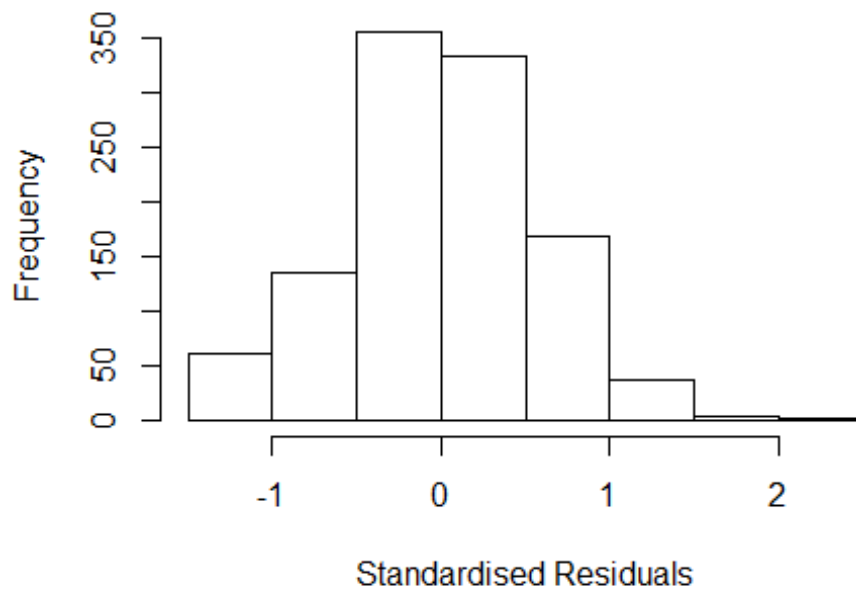
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3746 -0.3568 -0.0054 0.3982 2.2427
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3588 0.1344 10.11 <2e-16 ***
## FirstAuthorFemale1 0.0524 0.0402 1.30 0.192
## LastAuthorFemale1 -0.0366 0.0425 -0.86 0.389
## Year1997 -0.2490 0.1684 -1.48 0.139
## Year1998 -0.2083 0.1607 -1.30 0.195
## Year1999 -0.2889 0.1724 -1.68 0.094 .
## Year2000 -0.1926 0.1521 -1.27 0.206
## Year2001 -0.1355 0.1605 -0.84 0.399
## Year2002 -0.1939 0.1706 -1.14 0.256
## Year2003 -0.0784 0.1470 -0.53 0.594
## Year2004 -0.0666 0.1572 -0.42 0.672
## Year2005 -0.1028 0.1452 -0.71 0.479
```

```

## Year2006          -0.2021      0.1478   -1.37    0.172
## Year2007          -0.0570      0.1487   -0.38    0.702
## Year2008          -0.1689      0.1508   -1.12    0.263
## Year2009          -0.2436      0.1430   -1.70    0.089 .
## Year2010          -0.2088      0.1497   -1.40    0.163
## Year2011          -0.1211      0.1535   -0.79    0.430
## Year2012          -0.1770      0.1470   -1.20    0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.557
## Multiple R-squared:  0.0156, Adjusted R-squared:  -0.000872
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 1009 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0686 0.8740 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      9.12e-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 1.037 1      1.018
## Year              1.037 16      1.001

```

Residuals from first author

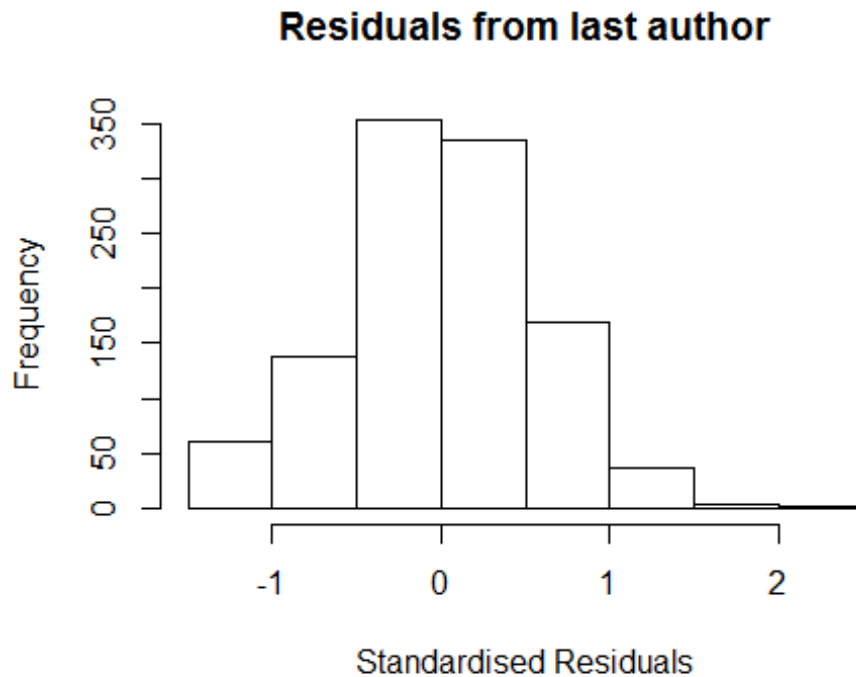


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38885 -0.35835 -0.00786  0.39453  2.25068
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3547    0.1340   10.11  <2e-16 ***
## FirstAuthorFemale1  0.0341    0.0351    0.97   0.332
## Year1997        -0.2496    0.1683   -1.48   0.138
## Year1998        -0.2105    0.1603   -1.31   0.190
## Year1999        -0.2955    0.1718   -1.72   0.086 .
## Year2000        -0.1893    0.1518   -1.25   0.212
## Year2001        -0.1394    0.1598   -0.87   0.383
## Year2002        -0.1960    0.1705   -1.15   0.250
## Year2003        -0.0789    0.1467   -0.54   0.591
## Year2004        -0.0640    0.1570   -0.41   0.684
## Year2005        -0.1039    0.1447   -0.72   0.473
## Year2006        -0.2010    0.1474   -1.36   0.173
```

```

## Year2007          -0.0573      0.1482   -0.39    0.699
## Year2008          -0.1682      0.1504   -1.12    0.263
## Year2009          -0.2430      0.1426   -1.70    0.089 .
## Year2010          -0.2096      0.1492   -1.40    0.160
## Year2011          -0.1235      0.1528   -0.81    0.419
## Year2012          -0.1795      0.1465   -1.23    0.221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0149, Adjusted R-squared:  -0.000588
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 1013 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0668 0.8730 0.9520 0.9060 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      9.12e-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 1.066 1      1.032
## Year      1.066 16      1.002

```

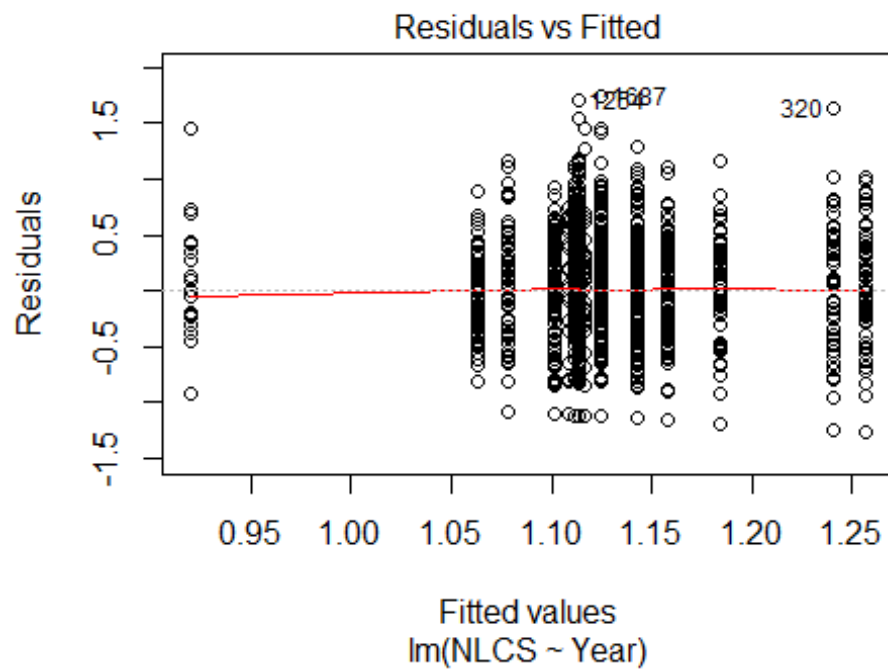
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36201 -0.36241 0.00115 0.39731 2.23851
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36201 0.13354 10.20 <2e-16 ***
## LastAuthorFemale1 -0.00839 0.03698 -0.23 0.821
## Year1997 -0.24525 0.16855 -1.46 0.146
## Year1998 -0.20574 0.16014 -1.28 0.199
## Year1999 -0.28591 0.17160 -1.67 0.096 .
## Year2000 -0.18756 0.15150 -1.24 0.216
## Year2001 -0.13452 0.16004 -0.84 0.401
## Year2002 -0.19021 0.16953 -1.12 0.262
## Year2003 -0.07072 0.14597 -0.48 0.628
## Year2004 -0.05594 0.15648 -0.36 0.721
## Year2005 -0.09383 0.14428 -0.65 0.516
## Year2006 -0.19470 0.14665 -1.33 0.185
```

```

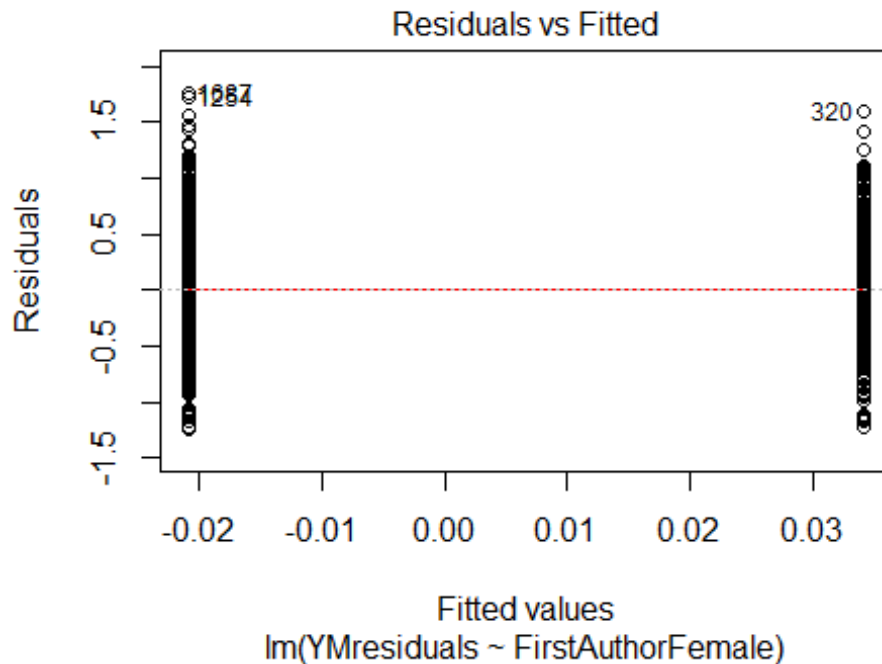
## Year2007          -0.04843      0.14749   -0.33      0.743
## Year2008          -0.15910      0.14964   -1.06      0.288
## Year2009          -0.23594      0.14190   -1.66      0.097 .
## Year2010          -0.19931      0.14896   -1.34      0.181
## Year2011          -0.11621      0.15288   -0.76      0.447
## Year2012          -0.16912      0.14619   -1.16      0.248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.557
## Multiple R-squared:  0.0142, Adjusted R-squared:  -0.00134
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 1016 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0695 0.8720 0.9520 0.9050 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      9.12e-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: 1096"
## [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
##   28   38   39   52   62   63   70   79   81   81  128  137  139  187  217
## 2011 2012
##   200   201
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   26   29   35   46   52   60   66   75   66  101  113  121  167  186
## 2011 2012

```

```
## 170 163
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 26 28 34 40 49 52 62 67 60 89 109 105 154 173
## 2011 2012
## 156 150
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```

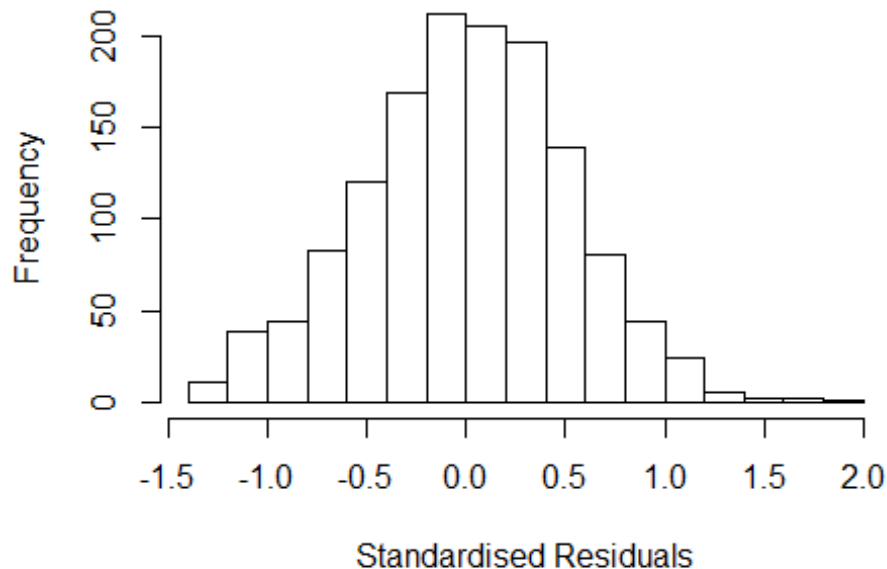


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.69, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.12507851548162"
## [1] "Male first author team size 2018 geometric mean: 2.10707492539104"
##
## 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: 2.05140313652869"
## [1] "Male last author team size 2018 geometric mean: 2.18213634775796"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, 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 1.405 1          1.185
## LastAuthorFemale  1.377 1          1.173
## UniqueAuthors    1.295 4          1.033
## Year              1.482 16         1.012
```

Residuals from first and last author and team size



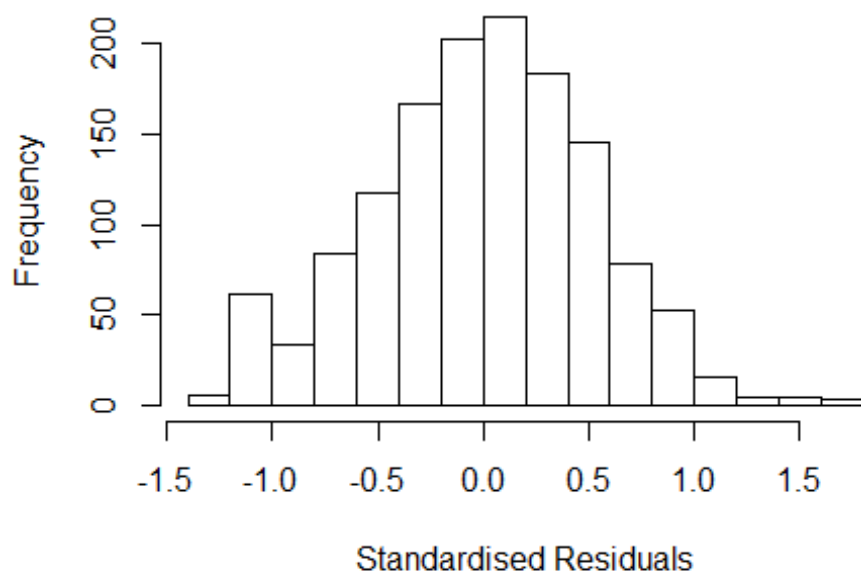
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30935 -0.34521  0.00817  0.34773  1.84524
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7899    0.1109   7.12 1.7e-12 ***
## FirstAuthorFemale1 0.0329    0.0336   0.98  0.328
## LastAuthorFemale1 0.0451    0.0333   1.36  0.175
## UniqueAuthors2    0.2045    0.0340   6.01 2.4e-09 ***
## UniqueAuthors3    0.2066    0.0412   5.01 6.0e-07 ***
## UniqueAuthors4    0.1006    0.0866   1.16  0.245
## UniqueAuthors5    0.1600    0.1051   1.52  0.128
## Year1997          0.2244    0.1709   1.31  0.189
## Year1998          0.2821    0.1596   1.77  0.077 .
## Year1999          0.2072    0.1524   1.36  0.174
```

```

## Year2000          0.2606      0.1336      1.95      0.051 .
## Year2001          0.3374      0.1366      2.47      0.014 *
## Year2002          0.3093      0.1354      2.28      0.023 *
## Year2003          0.1375      0.1313      1.05      0.295
## Year2004          0.1101      0.1220      0.90      0.367
## Year2005          0.2364      0.1251      1.89      0.059 .
## Year2006          0.2084      0.1225      1.70      0.089 .
## Year2007          0.1667      0.1215      1.37      0.170
## Year2008          0.1839      0.1212      1.52      0.130
## Year2009          0.1838      0.1204      1.53      0.127
## Year2010          0.1405      0.1208      1.16      0.245
## Year2011          0.1445      0.1212      1.19      0.233
## Year2012          0.1530      0.1223      1.25      0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.05,   Adjusted R-squared:  0.0345
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1265 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.156  0.868  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      7.27e-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 1.395 1      1.181
## LastAuthorFemale  1.378 1      1.174
## Year              1.187 16      1.005

```

Residuals from first and last author



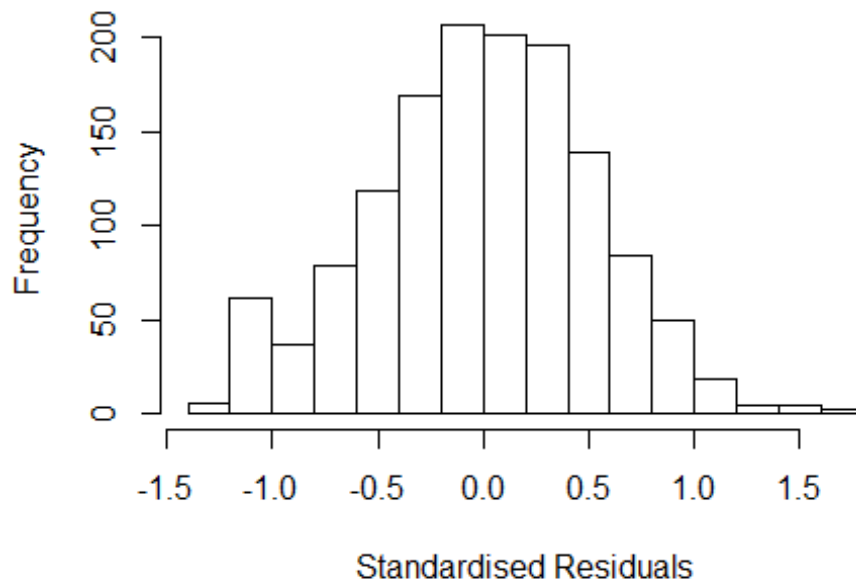
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2668 -0.3372 0.0134 0.3520 1.7424
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8691 0.1158 7.51 1.1e-13 ***
## FirstAuthorFemale1 0.0511 0.0342 1.50 0.135
## LastAuthorFemale1 0.0526 0.0340 1.55 0.122
## Year1997 0.2108 0.1723 1.22 0.221
## Year1998 0.2808 0.1662 1.69 0.091 .
## Year1999 0.2300 0.1523 1.51 0.131
## Year2000 0.2597 0.1389 1.87 0.062 .
## Year2001 0.3466 0.1442 2.40 0.016 *
## Year2002 0.3451 0.1401 2.46 0.014 *
## Year2003 0.1725 0.1358 1.27 0.204
## Year2004 0.1375 0.1269 1.08 0.279
## Year2005 0.2726 0.1310 2.08 0.038 *
```

```

## Year2006          0.2505      0.1278      1.96      0.050 .
## Year2007          0.2129      0.1265      1.68      0.093 .
## Year2008          0.2267      0.1267      1.79      0.074 .
## Year2009          0.2240      0.1257      1.78      0.075 .
## Year2010          0.1969      0.1245      1.58      0.114
## Year2011          0.1905      0.1263      1.51      0.132
## Year2012          0.2024      0.1270      1.59      0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0168, Adjusted R-squared:  0.00371
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1253 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.231  0.868  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      7.27e-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 1.112 1      1.054
## Year              1.112 16      1.003

```


Residuals from first author



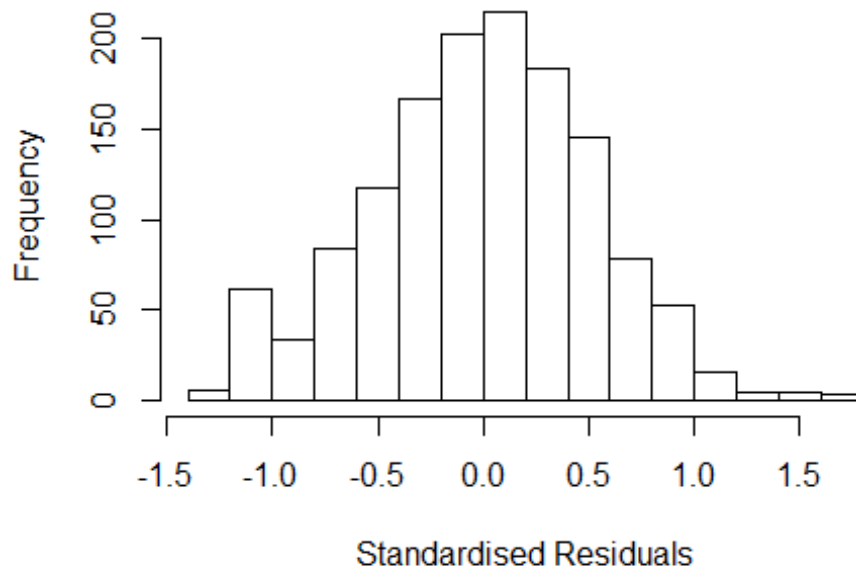
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2306 -0.3407 0.0106 0.3476 1.7834
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8671 0.1151 7.53 9.2e-14 ***
## FirstAuthorFemale1 0.0733 0.0305 2.41 0.016 *
## Year1997 0.2247 0.1724 1.30 0.193
## Year1998 0.2902 0.1654 1.75 0.080 .
## Year1999 0.2358 0.1524 1.55 0.122
## Year2000 0.2672 0.1381 1.94 0.053 .
## Year2001 0.3595 0.1438 2.50 0.013 *
## Year2002 0.3551 0.1394 2.55 0.011 *
## Year2003 0.1837 0.1355 1.36 0.176
## Year2004 0.1447 0.1261 1.15 0.251
## Year2005 0.2793 0.1304 2.14 0.032 *
## Year2006 0.2647 0.1264 2.09 0.037 *
```

```

## Year2007          0.2246      0.1255      1.79      0.074 .
## Year2008          0.2403      0.1255      1.91      0.056 .
## Year2009          0.2386      0.1246      1.92      0.056 .
## Year2010          0.2105      0.1234      1.71      0.088 .
## Year2011          0.2066      0.1250      1.65      0.099 .
## Year2012          0.2195      0.1255      1.75      0.080 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.00275
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.210  0.869  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      7.27e-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 1.092 1      1.045
## Year              1.092 16      1.003

```

Residuals from last author



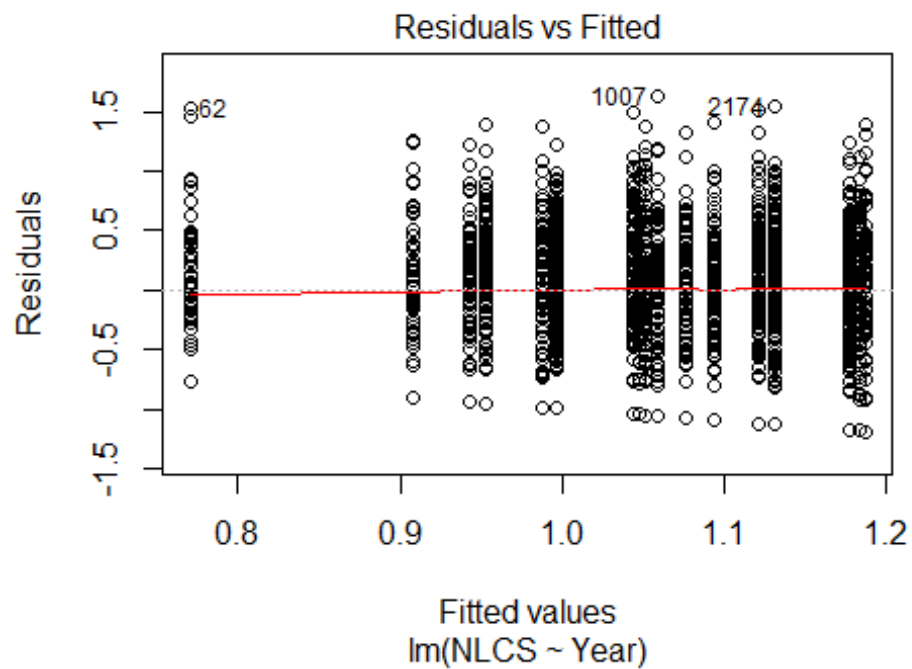
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30073 -0.33970 0.00947 0.36027 1.70947
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8750 0.1157 7.56 7.4e-14 ***
## LastAuthorFemale1 0.0748 0.0302 2.47 0.013 *
## Year1997 0.2131 0.1718 1.24 0.215
## Year1998 0.2840 0.1668 1.70 0.089 .
## Year1999 0.2250 0.1516 1.48 0.138
## Year2000 0.2627 0.1398 1.88 0.060 .
## Year2001 0.3529 0.1440 2.45 0.014 *
## Year2002 0.3509 0.1397 2.51 0.012 *
## Year2003 0.1795 0.1348 1.33 0.183
## Year2004 0.1460 0.1267 1.15 0.249
## Year2005 0.2787 0.1312 2.12 0.034 *
## Year2006 0.2526 0.1277 1.98 0.048 *
```

```

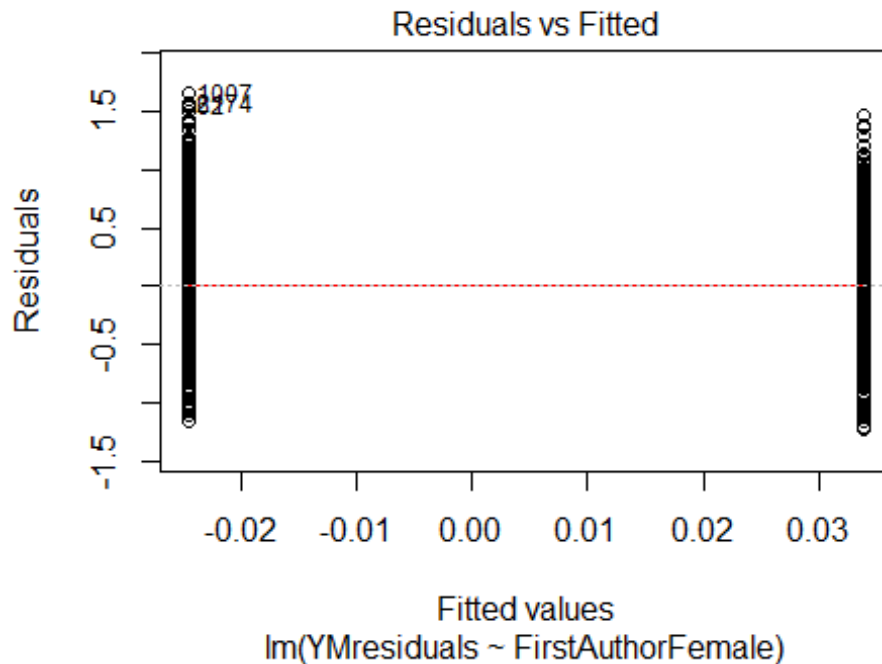
## Year2007          0.2196      0.1264      1.74      0.083 .
## Year2008          0.2379      0.1264      1.88      0.060 .
## Year2009          0.2308      0.1255      1.84      0.066 .
## Year2010          0.2017      0.1244      1.62      0.105
## Year2011          0.1966      0.1262      1.56      0.119
## Year2012          0.2066      0.1270      1.63      0.104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00267
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.249  0.869  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      7.27e-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: 1375"
## [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
##   79   83   79   90   84  110  109  101  106  125  124  144  151  182  204
## 2011 2012
##   201  210
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   72   76   64   83   69   97   94   87   93  109  109  127  137  158  176
## 2011 2012

```

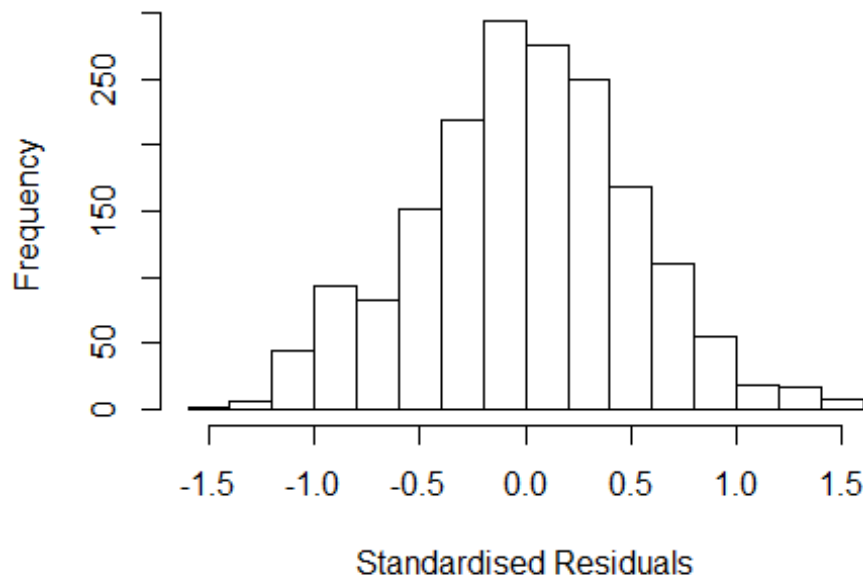
```
## 175 185
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 72 63 80 66 92 87 83 88 108 102 123 126 146 162
## 2011 2012
## 160 168
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 10, df = 16, p-value = 0.9
```



```
##
## 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



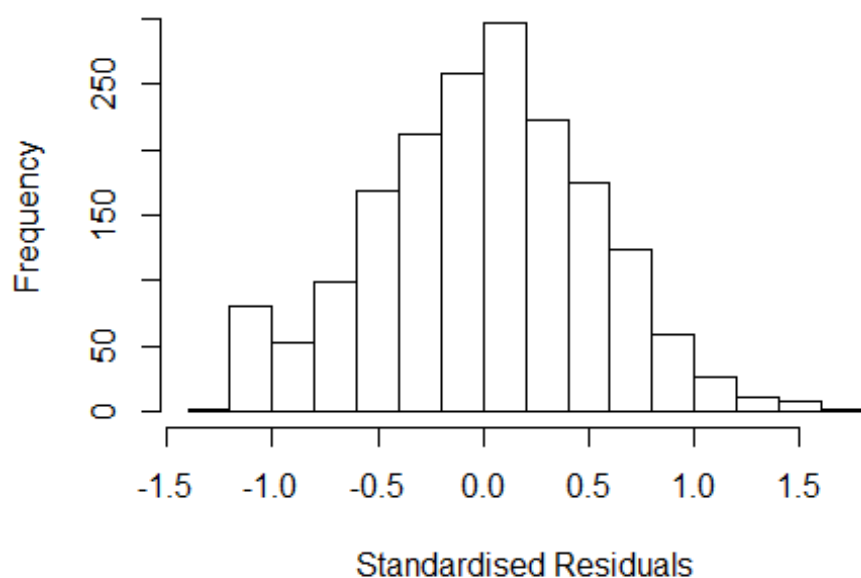
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40502 -0.32922 0.00161 0.33850 1.59072
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.65471 0.06144 10.66 < 2e-16 ***
## FirstAuthorFemale1 0.04506 0.02818 1.60 0.10997
## LastAuthorFemale1 0.00937 0.02811 0.33 0.73909
## UniqueAuthors2 0.17920 0.02887 6.21 6.7e-10 ***
## UniqueAuthors3 0.22484 0.03793 5.93 3.7e-09 ***
## UniqueAuthors4 0.36030 0.05341 6.75 2.1e-11 ***
## UniqueAuthors5 0.39246 0.17150 2.29 0.02223 *
## Year1997 0.18951 0.08455 2.24 0.02512 *
## Year1998 0.11747 0.09713 1.21 0.22668
## Year1999 0.23718 0.09463 2.51 0.01228 *
```

```

## Year2000          0.31051      0.08929      3.48  0.00052 ***
## Year2001          0.26312      0.08306      3.17  0.00156 **
## Year2002          0.38749      0.08449      4.59  4.8e-06 ***
## Year2003          0.42024      0.08642      4.86  1.3e-06 ***
## Year2004          0.30035      0.08527      3.52  0.00044 ***
## Year2005          0.26336      0.08024      3.28  0.00105 **
## Year2006          0.28498      0.07677      3.71  0.00021 ***
## Year2007          0.24462      0.07876      3.11  0.00193 **
## Year2008          0.14916      0.07265      2.05  0.04020 *
## Year2009          0.35993      0.07533      4.78  1.9e-06 ***
## Year2010          0.18461      0.07669      2.41  0.01617 *
## Year2011          0.30343      0.07620      3.98  7.1e-05 ***
## Year2012          0.29984      0.07228      4.15  3.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0801, Adjusted R-squared:  0.0687
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1643 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.296  0.865   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      5.58e-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 1.297 1      1.139
## LastAuthorFemale  1.247 1      1.117
## Year              1.088 16      1.003

```


Residuals from first and last author



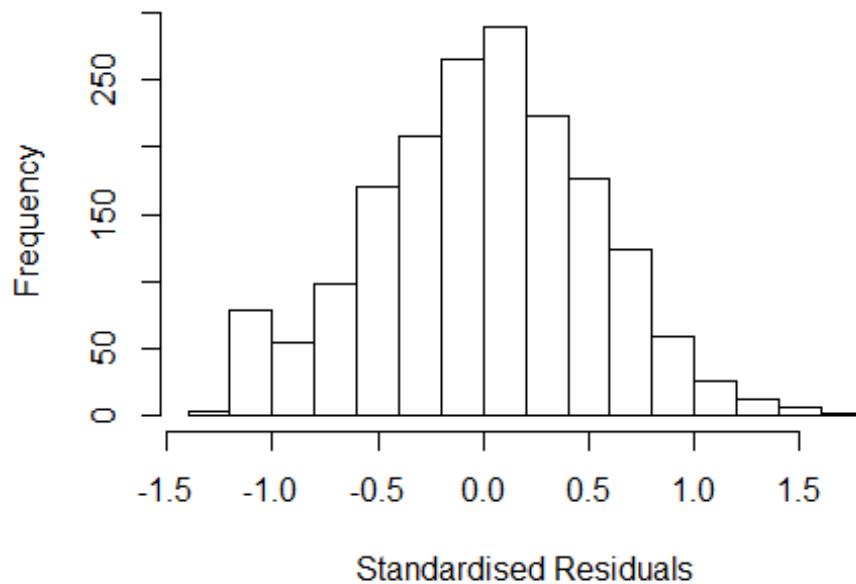
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2070 -0.3423 0.0126 0.3483 1.6587
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7412 0.0622 11.92 < 2e-16 ***
## FirstAuthorFemale1 0.0561 0.0291 1.93 0.05411 .
## LastAuthorFemale1 0.0090 0.0290 0.31 0.75623
## Year1997 0.1795 0.0860 2.09 0.03700 *
## Year1998 0.1181 0.0956 1.23 0.21711
## Year1999 0.2262 0.0945 2.39 0.01682 *
## Year2000 0.2831 0.0892 3.18 0.00152 **
## Year2001 0.2658 0.0824 3.23 0.00127 **
## Year2002 0.4007 0.0864 4.64 3.8e-06 ***
## Year2003 0.4249 0.0873 4.87 1.2e-06 ***
## Year2004 0.3106 0.0854 3.64 0.00028 ***
## Year2005 0.2881 0.0812 3.55 0.00040 ***
```

```

## Year2006          0.3201      0.0787      4.07  4.9e-05 ***
## Year2007          0.2953      0.0804      3.67  0.00025 ***
## Year2008          0.1477      0.0757      1.95  0.05119 .
## Year2009          0.3924      0.0759      5.17  2.6e-07 ***
## Year2010          0.2173      0.0779      2.79  0.00533 **
## Year2011          0.3468      0.0773      4.49  7.6e-06 ***
## Year2012          0.3343      0.0739      4.52  6.5e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0385, Adjusted R-squared:  0.0287
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1625 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.285  0.869  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      5.58e-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 1.067 1      1.033
## Year      1.067 16      1.002

```

Residuals from first author



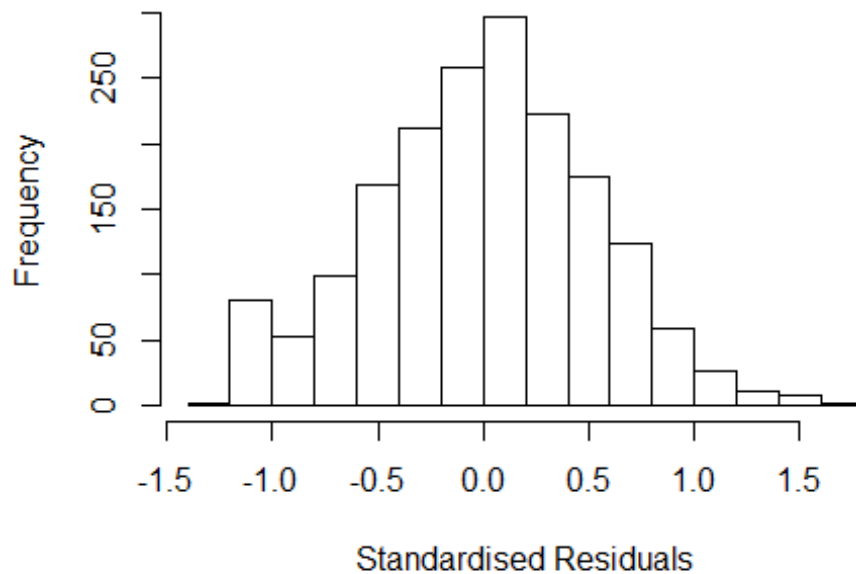
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2043 -0.3411 0.0118 0.3514 1.6566
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7422 0.0622 11.94 < 2e-16 ***
## FirstAuthorFemale1 0.0607 0.0265 2.29 0.02229 *
## Year1997 0.1805 0.0859 2.10 0.03573 *
## Year1998 0.1184 0.0956 1.24 0.21571
## Year1999 0.2265 0.0943 2.40 0.01647 *
## Year2000 0.2833 0.0891 3.18 0.00149 **
## Year2001 0.2662 0.0823 3.24 0.00124 **
## Year2002 0.4014 0.0862 4.66 3.5e-06 ***
## Year2003 0.4253 0.0872 4.88 1.2e-06 ***
## Year2004 0.3109 0.0853 3.64 0.00028 ***
## Year2005 0.2892 0.0811 3.57 0.00037 ***
## Year2006 0.3212 0.0785 4.09 4.5e-05 ***
```

```

## Year2007          0.2966      0.0801      3.70  0.00022 ***
## Year2008          0.1484      0.0755      1.97  0.04951 *
## Year2009          0.3927      0.0759      5.18  2.5e-07 ***
## Year2010          0.2180      0.0777      2.80  0.00510 **
## Year2011          0.3479      0.0771      4.51  6.8e-06 ***
## Year2012          0.3347      0.0738      4.53  6.2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.0291
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1623 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.288  0.868  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      5.58e-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 1.028 1      1.014
## Year              1.028 16      1.001

```

Residuals from last author



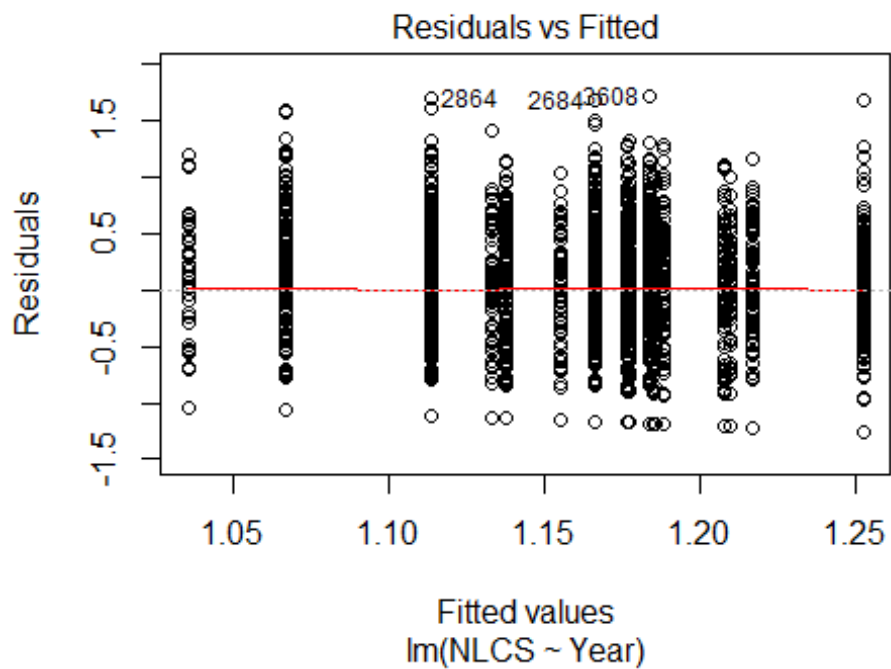
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2158 -0.3436 0.0172 0.3565 1.6442
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7493 0.0621 12.07 < 2e-16 ***
## LastAuthorFemale1 0.0377 0.0264 1.43 0.15294
## Year1997 0.1755 0.0858 2.05 0.04095 *
## Year1998 0.1224 0.0956 1.28 0.20090
## Year1999 0.2269 0.0945 2.40 0.01649 *
## Year2000 0.2832 0.0895 3.17 0.00157 **
## Year2001 0.2688 0.0826 3.25 0.00116 **
## Year2002 0.4062 0.0866 4.69 2.9e-06 ***
## Year2003 0.4288 0.0874 4.90 1.0e-06 ***
## Year2004 0.3177 0.0859 3.70 0.00022 ***
## Year2005 0.2944 0.0812 3.63 0.00030 ***
## Year2006 0.3256 0.0785 4.15 3.5e-05 ***
```

```

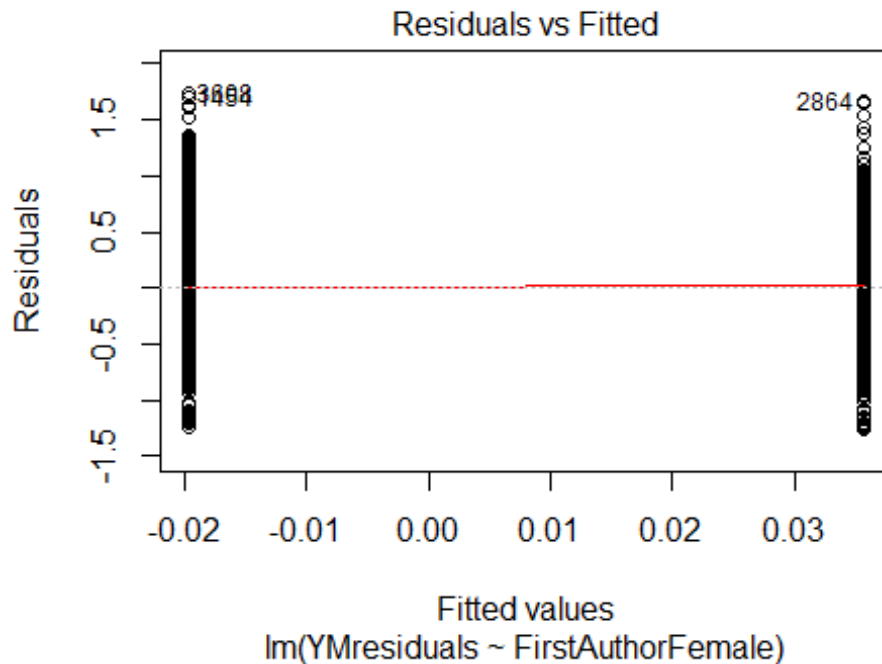
## Year2007          0.2946      0.0804      3.67  0.00025 ***
## Year2008          0.1525      0.0759      2.01  0.04463 *
## Year2009          0.4045      0.0753      5.37  8.9e-08 ***
## Year2010          0.2183      0.0780      2.80  0.00518 **
## Year2011          0.3515      0.0773      4.55  5.8e-06 ***
## Year2012          0.3401      0.0740      4.60  4.6e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0366, Adjusted R-squared:  0.0274
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1632 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.297  0.870  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      5.58e-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: 1793"
## [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
##   60   72   80  109  127  139  128  121  101  163  156  217  225  280  292
## 2011 2012
##  292  329
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   49   61   79   82   97   97   94   84  134  125  166  182  221  243
## 2011 2012

```

```
## 234 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 44 58 70 77 89 87 83 81 128 113 150 159 195 222
## 2011 2012
## 212 226
## [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
```

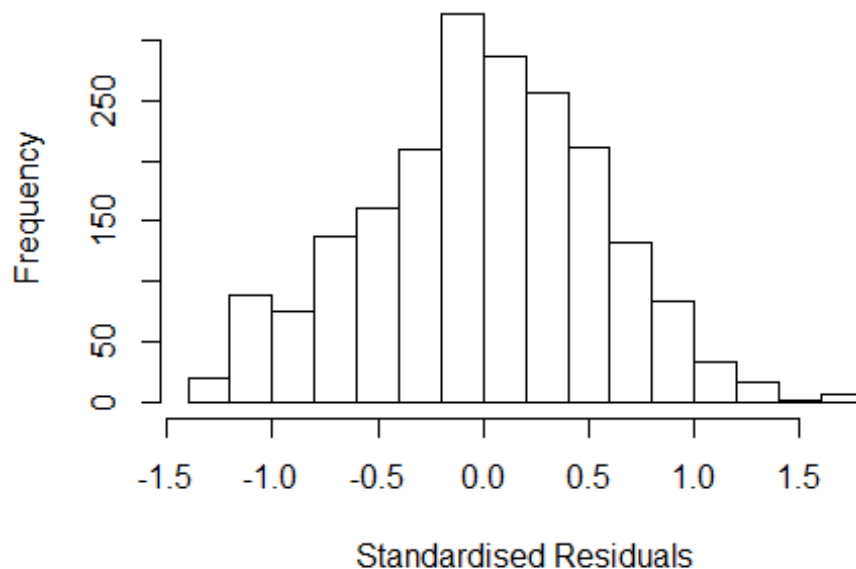


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.39, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 2.06956707186619"
## [1] "Male first author team size 2018 geometric mean: 2.34683483860169"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4600, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.20826354424544"
## [1] "Male last author team size 2018 geometric mean: 2.25908724510525"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, 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 1.415 1          1.189
## LastAuthorFemale  1.383 1          1.176
## UniqueAuthors    1.216 4          1.025
## Year              1.278 16         1.008
```


Residuals from first and last author and team size



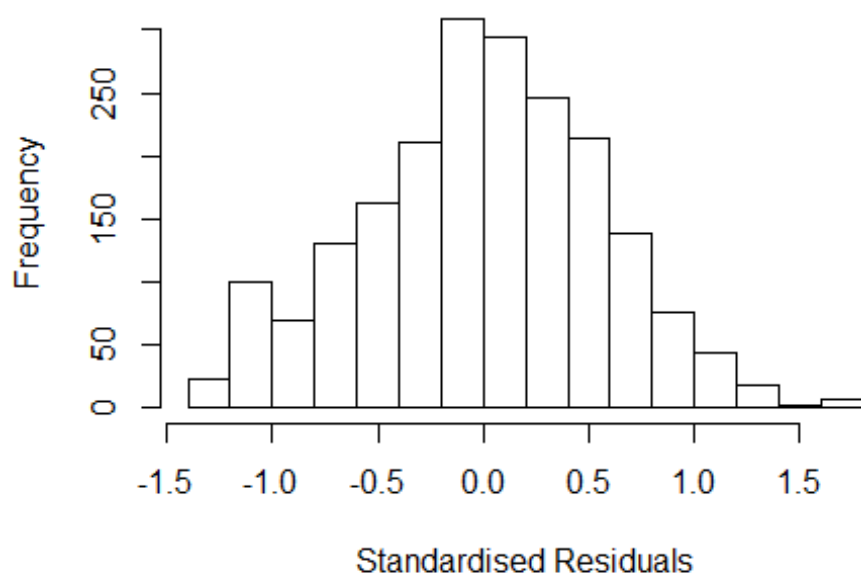
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37696 -0.36945  0.00669  0.38503  1.76060
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9732    0.0973   10.00 < 2e-16 ***
## FirstAuthorFemale1  0.0834    0.0307    2.71  0.0067 **
## LastAuthorFemale1 -0.0424    0.0324   -1.31  0.1911
## UniqueAuthors2    0.0931    0.0296    3.15  0.0017 **
## UniqueAuthors3    0.2173    0.0377    5.76 9.9e-09 ***
## UniqueAuthors4    0.1727    0.0660    2.62  0.0089 **
## UniqueAuthors5    0.0251    0.1166    0.22  0.8296
## Year1997          0.1306    0.1186    1.10  0.2709
## Year1998          0.0544    0.1201    0.45  0.6506
## Year1999          0.0931    0.1150    0.81  0.4180
```

```

## Year2000          0.1512      0.1108      1.36      0.1726
## Year2001          0.1786      0.1149      1.55      0.1203
## Year2002          0.1443      0.1206      1.20      0.2317
## Year2003          0.1851      0.1136      1.63      0.1034
## Year2004          0.1751      0.1103      1.59      0.1127
## Year2005          0.1338      0.1082      1.24      0.2162
## Year2006          0.2012      0.1075      1.87      0.0614 .
## Year2007          0.1030      0.1058      0.97      0.3303
## Year2008          0.0891      0.1066      0.84      0.4031
## Year2009         -0.0194      0.1081     -0.18      0.8578
## Year2010          0.0872      0.1072      0.81      0.4158
## Year2011          0.0294      0.1056      0.28      0.7805
## Year2012          0.0907      0.1045      0.87      0.3854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0325, Adjusted R-squared:  0.022
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 182 weights are ~= 1. The remaining 1859 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.299  0.865  0.950  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      4.90e-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 1.386 1      1.177
## LastAuthorFemale  1.374 1      1.172
## Year              1.090 16      1.003

```

Residuals from first and last author



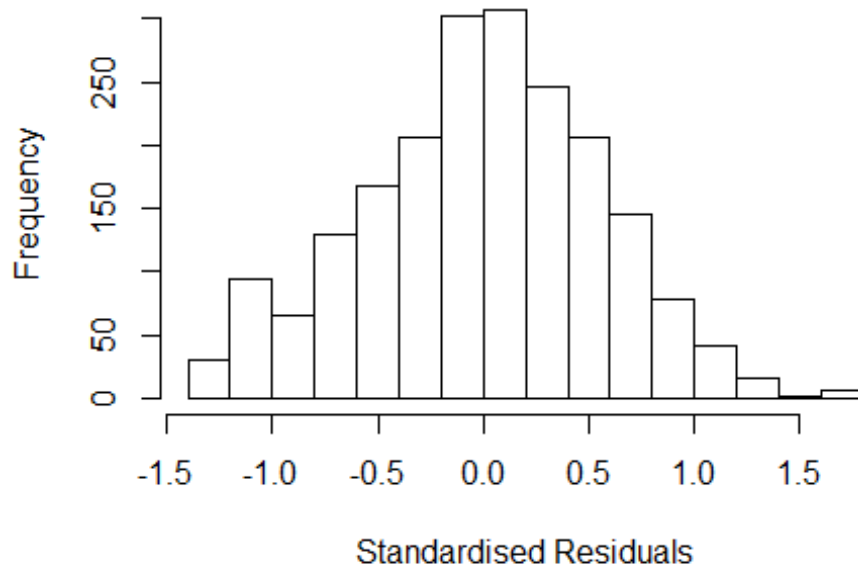
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2602 -0.3744  0.0125  0.3826  1.7366
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0332     0.0953   10.84  <2e-16 ***
## FirstAuthorFemale1  0.0967     0.0309    3.12  0.0018 **
## LastAuthorFemale1 -0.0520     0.0328   -1.58  0.1131
## Year1997          0.1336     0.1170    1.14  0.2536
## Year1998          0.0560     0.1191    0.47  0.6380
## Year1999          0.0984     0.1135    0.87  0.3864
## Year2000          0.1371     0.1102    1.24  0.2137
## Year2001          0.1713     0.1138    1.51  0.1323
## Year2002          0.1529     0.1186    1.29  0.1977
## Year2003          0.1824     0.1125    1.62  0.1053
## Year2004          0.1653     0.1098    1.51  0.1324
## Year2005          0.1282     0.1071    1.20  0.2316
```

```

## Year2006          0.2114      0.1060      1.99      0.0462 *
## Year2007          0.1137      0.1045      1.09      0.2766
## Year2008          0.0972      0.1056      0.92      0.3575
## Year2009         -0.0122      0.1075     -0.11      0.9097
## Year2010          0.1172      0.1057      1.11      0.2676
## Year2011          0.0574      0.1048      0.55      0.5842
## Year2012          0.1282      0.1030      1.24      0.2136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0147, Adjusted R-squared:  0.00593
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 1872 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.316  0.863   0.949   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      4.90e-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 1.047 1          1.023
## Year              1.047 16          1.001

```

Residuals from first author



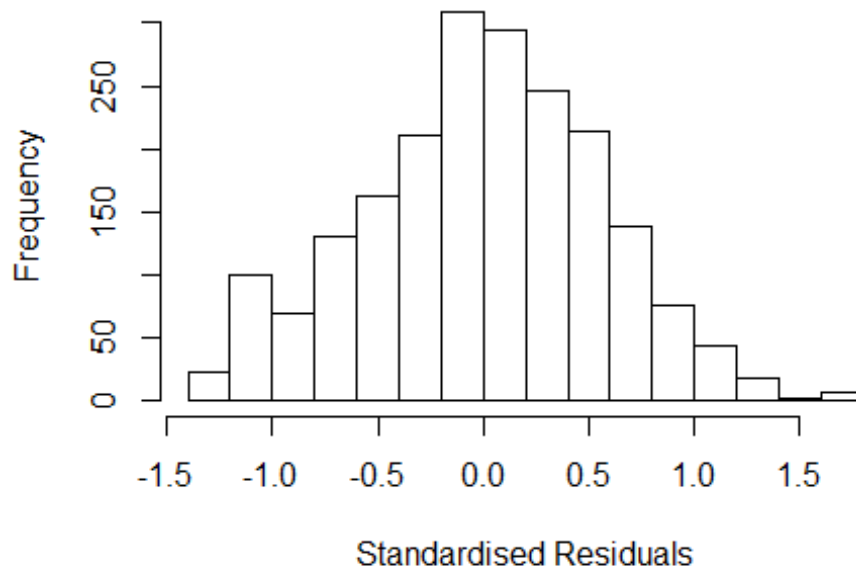
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2788 -0.3731 0.0144 0.3865 1.7454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0309 0.0954 10.81 <2e-16 ***
## FirstAuthorFemale1 0.0709 0.0270 2.62 0.0087 **
## Year1997 0.1271 0.1174 1.08 0.2790
## Year1998 0.0506 0.1189 0.43 0.6705
## Year1999 0.0934 0.1134 0.82 0.4105
## Year2000 0.1370 0.1103 1.24 0.2144
## Year2001 0.1646 0.1137 1.45 0.1478
## Year2002 0.1493 0.1188 1.26 0.2092
## Year2003 0.1770 0.1129 1.57 0.1170
## Year2004 0.1611 0.1099 1.47 0.1427
## Year2005 0.1233 0.1071 1.15 0.2497
## Year2006 0.2072 0.1059 1.96 0.0504 .
```

```

## Year2007          0.1105      0.1044      1.06      0.2900
## Year2008          0.0954      0.1058      0.90      0.3674
## Year2009         -0.0161      0.1076     -0.15      0.8815
## Year2010          0.1098      0.1055      1.04      0.2980
## Year2011          0.0523      0.1047      0.50      0.6172
## Year2012          0.1217      0.1029      1.18      0.2371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0134, Adjusted R-squared:  0.00514
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1867 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.311  0.863   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      4.90e-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 1.039 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
## -1.2614 -0.3805 0.0166 0.3856 1.7178
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.043679 0.095479 10.93 <2e-16 ***
## LastAuthorFemale1 -0.000592 0.028467 -0.02 0.983
## Year1997 0.133295 0.118602 1.12 0.261
## Year1998 0.057190 0.119044 0.48 0.631
## Year1999 0.098658 0.114192 0.86 0.388
## Year2000 0.143269 0.110599 1.30 0.195
## Year2001 0.173821 0.114001 1.52 0.127
## Year2002 0.158452 0.118511 1.34 0.181
## Year2003 0.200051 0.112743 1.77 0.076 .
## Year2004 0.175310 0.110190 1.59 0.112
## Year2005 0.137078 0.107605 1.27 0.203
## Year2006 0.217749 0.106155 2.05 0.040 *
```

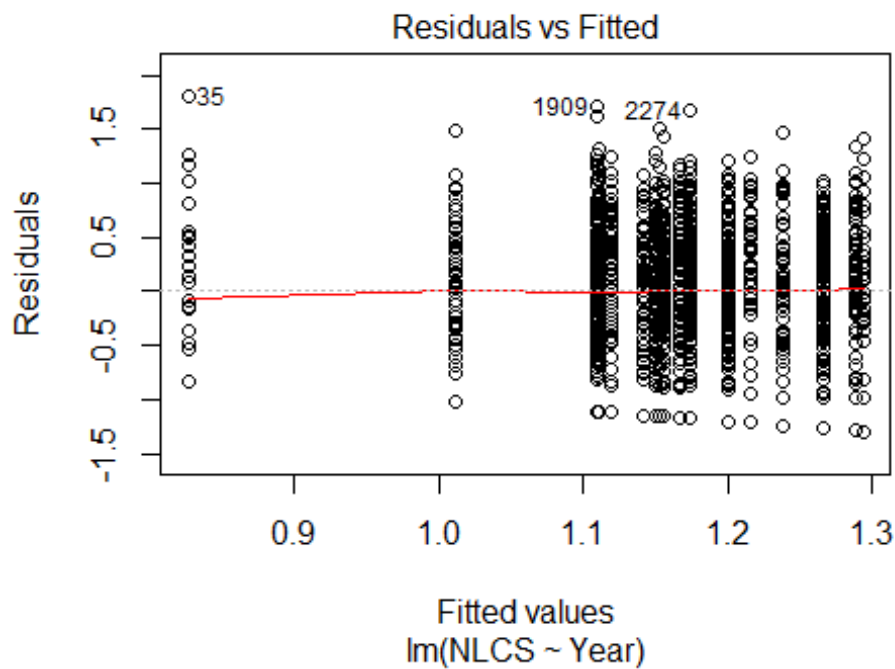
```

## Year2007      0.120896  0.104848  1.15  0.249
## Year2008      0.111972  0.106052  1.06  0.291
## Year2009      0.003977  0.107534  0.04  0.971
## Year2010      0.124968  0.105888  1.18  0.238
## Year2011      0.065554  0.105211  0.62  0.533
## Year2012      0.136492  0.103388  1.32  0.187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.00984,    Adjusted R-squared:  0.00152
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 1856 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.331  0.863  0.948  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      4.90e-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: 2041"
## [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
##   45   52   54   69   77  107  119  110   80   99  141  143  168  197  200
## 2011 2012
##   204  222
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   37   43   55   61   78   95   92   67   83  116  117  137  178  169
## 2011 2012

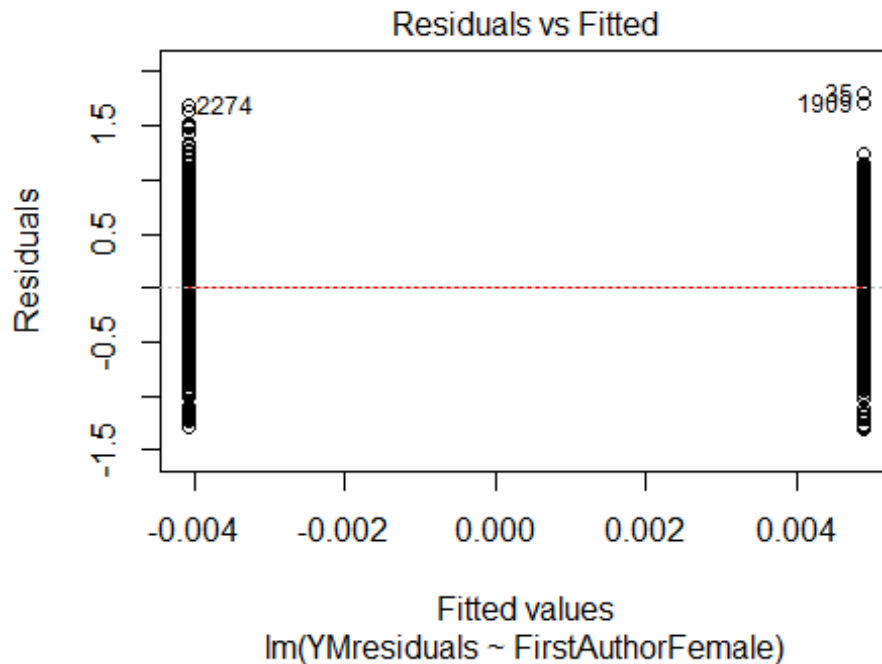
```



```
## 162 177
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 31 34 41 51 55 68 83 87 63 74 110 110 124 165 156
## 2011 2012
## 156 162
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.003
```

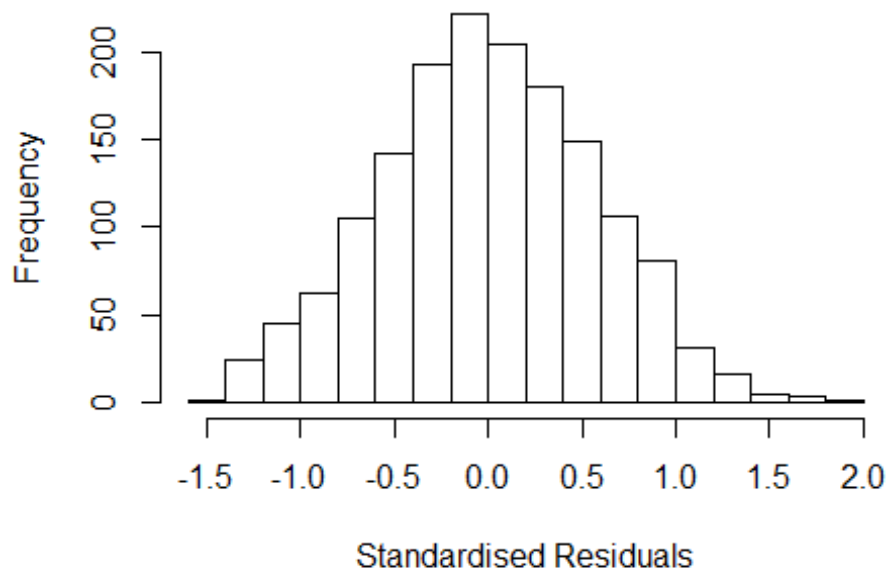


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.7854673852561"
## [1] "Male first author team size 2018 geometric mean: 2.10654851366937"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.81269809328028"
## [1] "Male last author team size 2018 geometric mean: 2.06058652473597"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.245 1      1.116
## LastAuthorFemale  1.238 1      1.113
## UniqueAuthors    1.278 4      1.031
## Year              1.396 16     1.010
```

Residuals from first and last author and team size



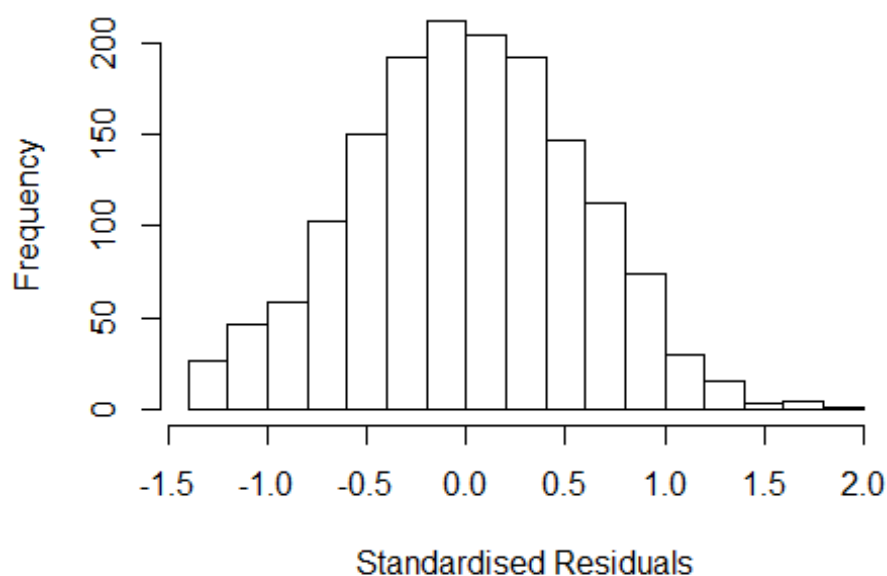
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4456 -0.3874 -0.0073 0.3957 1.8872
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7320 0.1414 5.18 2.6e-07 ***
## FirstAuthorFemale1 0.0607 0.0326 1.86 0.06285 .
## LastAuthorFemale1 -0.0509 0.0331 -1.54 0.12438
## UniqueAuthors2 0.0420 0.0348 1.20 0.22864
## UniqueAuthors3 0.0975 0.0453 2.15 0.03152 *
## UniqueAuthors4 0.0569 0.0873 0.65 0.51441
## UniqueAuthors5 0.0811 0.1081 0.75 0.45326
## Year1997 0.3546 0.1821 1.95 0.05164 .
## Year1998 0.5209 0.1887 2.76 0.00584 **
## Year1999 0.4579 0.1870 2.45 0.01442 *
```

```

## Year2000          0.4752      0.1667      2.85  0.00441 **
## Year2001          0.5555      0.1632      3.40  0.00068 ***
## Year2002          0.4535      0.1576      2.88  0.00405 **
## Year2003          0.4026      0.1595      2.52  0.01171 *
## Year2004          0.3580      0.1628      2.20  0.02799 *
## Year2005          0.2612      0.1571      1.66  0.09646 .
## Year2006          0.3479      0.1510      2.30  0.02139 *
## Year2007          0.3781      0.1507      2.51  0.01224 *
## Year2008          0.3694      0.1516      2.44  0.01491 *
## Year2009          0.4187      0.1494      2.80  0.00513 **
## Year2010          0.4736      0.1499      3.16  0.00161 **
## Year2011          0.3161      0.1505      2.10  0.03591 *
## Year2012          0.3843      0.1514      2.54  0.01125 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.0287, Adjusted R-squared:  0.0149
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1441 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.872  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      6.37e-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 1.236 1      1.112
## LastAuthorFemale  1.213 1      1.101
## Year              1.117 16      1.003

```

Residuals from first and last author



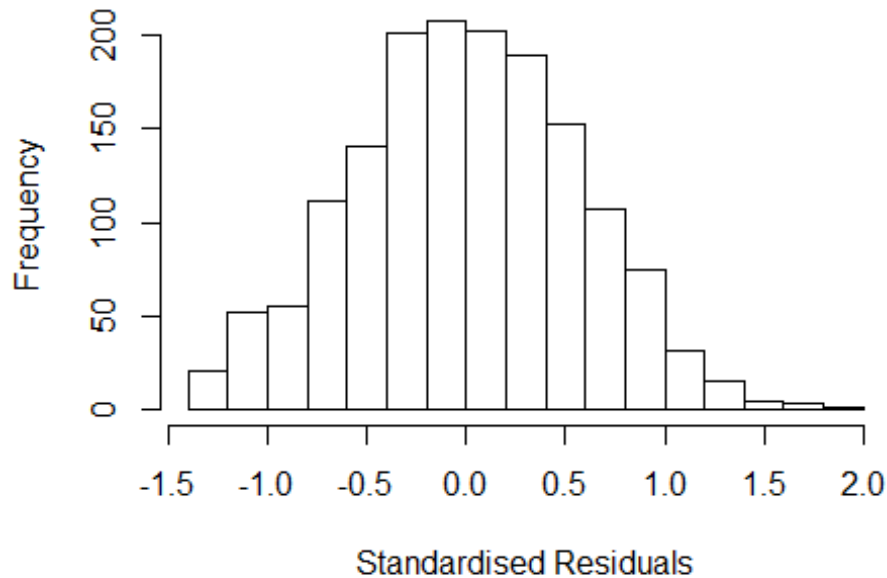
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3833 -0.3878 -0.0027 0.3961 1.8806
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7399 0.1421 5.21 2.2e-07 ***
## FirstAuthorFemale1 0.0621 0.0326 1.90 0.05711 .
## LastAuthorFemale1 -0.0536 0.0329 -1.63 0.10299
## Year1997 0.3615 0.1814 1.99 0.04642 *
## Year1998 0.5405 0.1888 2.86 0.00426 **
## Year1999 0.4723 0.1863 2.54 0.01132 *
## Year2000 0.4874 0.1673 2.91 0.00363 **
## Year2001 0.5813 0.1624 3.58 0.00036 ***
## Year2002 0.4633 0.1583 2.93 0.00348 **
## Year2003 0.4156 0.1600 2.60 0.00947 **
## Year2004 0.3795 0.1631 2.33 0.02009 *
## Year2005 0.2790 0.1571 1.78 0.07597 .
```

```

## Year2006          0.3673      0.1507      2.44  0.01492 *
## Year2007          0.4110      0.1504      2.73  0.00636 **
## Year2008          0.3972      0.1508      2.63  0.00852 **
## Year2009          0.4448      0.1490      2.98  0.00289 **
## Year2010          0.5067      0.1488      3.41  0.00068 ***
## Year2011          0.3534      0.1493      2.37  0.01806 *
## Year2012          0.4198      0.1500      2.80  0.00519 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0254, Adjusted R-squared:  0.014
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1439 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.870   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      6.37e-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 1.085 1      1.042
## Year              1.085 16      1.003

```

Residuals from first author



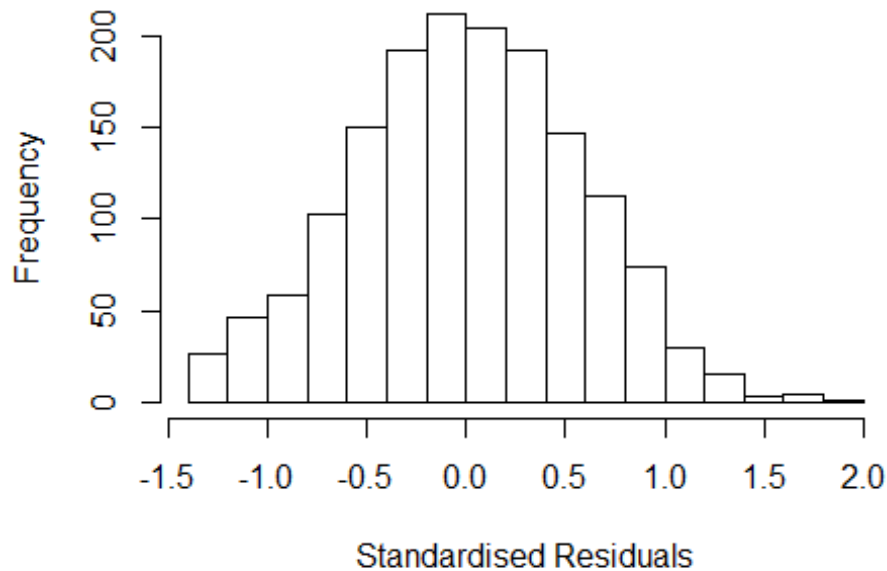
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34968 -0.38504 -0.00499  0.39637  1.85075
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7404     0.1427   5.19 2.4e-07 ***
## FirstAuthorFemale1 0.0379     0.0308   1.23 0.21833
## Year1997        0.3548     0.1814   1.96 0.05071 .
## Year1998        0.5341     0.1895   2.82 0.00488 **
## Year1999        0.4675     0.1868   2.50 0.01243 *
## Year2000        0.4749     0.1674   2.84 0.00462 **
## Year2001        0.5714     0.1631   3.50 0.00047 ***
## Year2002        0.4552     0.1588   2.87 0.00420 **
## Year2003        0.4069     0.1606   2.53 0.01138 *
## Year2004        0.3656     0.1636   2.23 0.02560 *
## Year2005        0.2661     0.1574   1.69 0.09117 .
## Year2006        0.3564     0.1511   2.36 0.01850 *
```

```

## Year2007          0.4021      0.1508      2.67  0.00775 **
## Year2008          0.3842      0.1512      2.54  0.01113 *
## Year2009          0.4313      0.1492      2.89  0.00390 **
## Year2010          0.4931      0.1491      3.31  0.00097 ***
## Year2011          0.3375      0.1496      2.26  0.02416 *
## Year2012          0.4061      0.1502      2.70  0.00694 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.285  0.872  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      6.37e-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 1.07 1      1.034
## Year              1.07 16      1.002

```


Residuals from last author



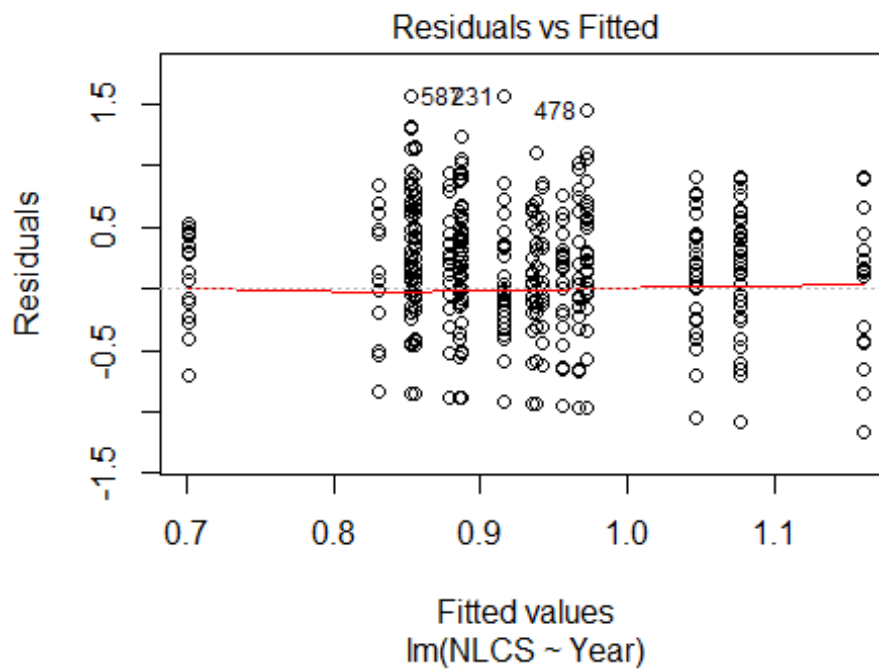
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33766 -0.39587 -0.00664 0.39299 1.90101
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7529 0.1407 5.35 1e-07 ***
## LastAuthorFemale1 -0.0249 0.0310 -0.80 0.42138
## Year1997 0.3549 0.1807 1.96 0.04969 *
## Year1998 0.5397 0.1873 2.88 0.00401 **
## Year1999 0.4698 0.1857 2.53 0.01152 *
## Year2000 0.4834 0.1665 2.90 0.00374 **
## Year2001 0.5847 0.1616 3.62 0.00031 ***
## Year2002 0.4638 0.1572 2.95 0.00323 **
## Year2003 0.4201 0.1591 2.64 0.00836 **
## Year2004 0.3828 0.1619 2.36 0.01817 *
## Year2005 0.2807 0.1561 1.80 0.07230 .
## Year2006 0.3709 0.1496 2.48 0.01327 *
```

```

## Year2007          0.4177      0.1493      2.80  0.00522 **
## Year2008          0.4040      0.1497      2.70  0.00703 **
## Year2009          0.4539      0.1477      3.07  0.00216 **
## Year2010          0.5117      0.1477      3.46  0.00055 ***
## Year2011          0.3574      0.1481      2.41  0.01593 *
## Year2012          0.4242      0.1488      2.85  0.00443 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0231, Adjusted R-squared:  0.0124
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1440 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.259  0.872  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      6.37e-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: 1570"
## [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
##   27   23   30   40   33   28   37   29   29   17   23   40   57   57   62
## 2011 2012
##   61   56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   19   29   32   28   23   33   26   26   15   23   35   51   52   59
## 2011 2012

```

```
## 56 52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 19 29 29 28 22 33 26 24 15 22 33 49 49 59
## 2011 2012
## 54 48
## [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
```



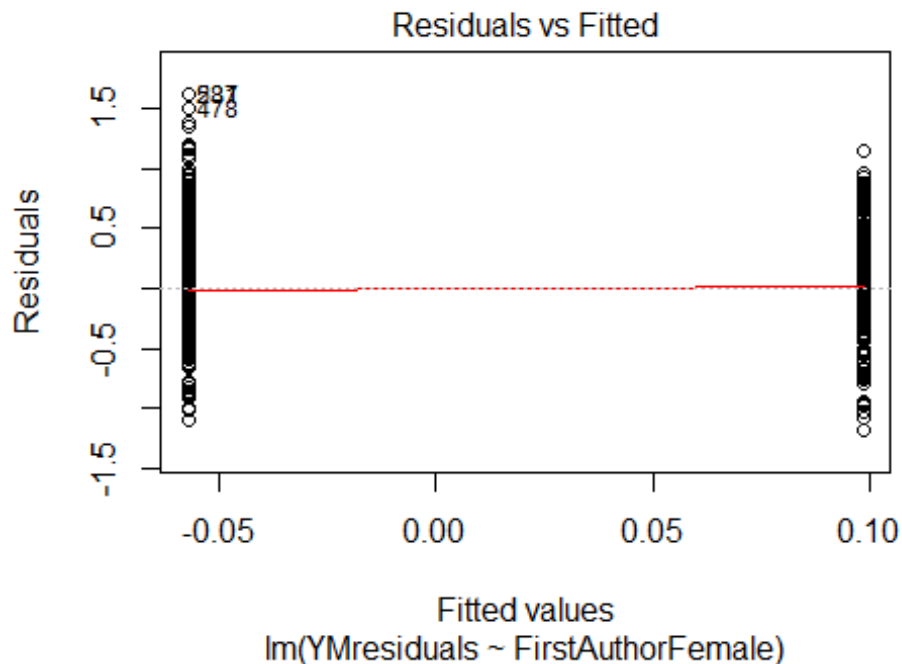
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 1.79407576626358"
## [1] "Male first author team size 2018 geometric mean: 1.52781848131706"

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

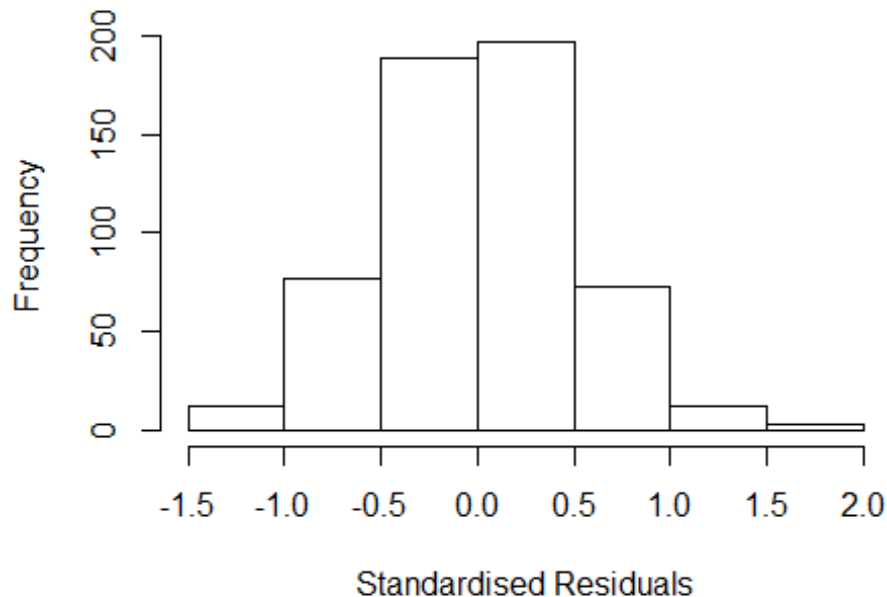
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.59488553429054"
## [1] "Male last author team size 2018 geometric mean: 1.63379241069886"

## 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 = 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 2.199  1      1.483
## LastAuthorFemale  1.990  1      1.411
## UniqueAuthors    1.767  4      1.074
## Year              2.175 16      1.025
```

Residuals from first and last author and team size



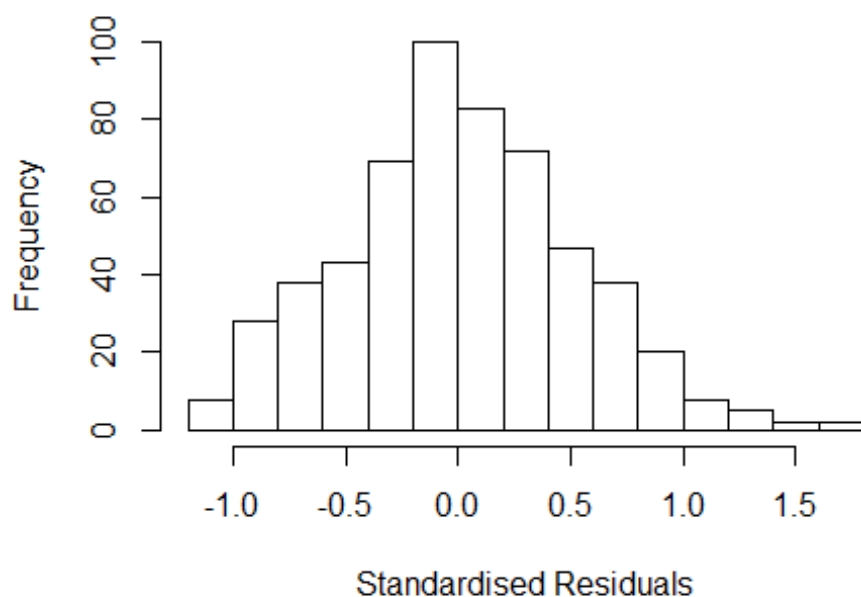
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.328383 -0.318957 0.000446 0.305442 1.767847
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8307 0.1113 7.46 3.4e-13 ***
## FirstAuthorFemale1 0.1527 0.0615 2.48 0.013 *
## LastAuthorFemale1 0.0458 0.0605 0.76 0.449
## UniqueAuthors2 0.1157 0.0506 2.29 0.023 *
## UniqueAuthors3 0.3938 0.0858 4.59 5.5e-06 ***
## UniqueAuthors4 0.2987 0.1453 2.06 0.040 *
## UniqueAuthors5 0.3228 0.4089 0.79 0.430
## Year1997 0.0598 0.1514 0.39 0.693
## Year1998 -0.0581 0.1375 -0.42 0.673
## Year1999 -0.0118 0.1489 -0.08 0.937
```

```

## Year2000          -0.2976      0.1327    -2.24      0.025 *
## Year2001          -0.0610      0.1430     -0.43      0.670
## Year2002          -0.1215      0.1312     -0.93      0.355
## Year2003          -0.0520      0.1504     -0.35      0.730
## Year2004          -0.0715      0.1420     -0.50      0.615
## Year2005          -0.1933      0.1768     -1.09      0.275
## Year2006           0.1039      0.1464      0.71      0.478
## Year2007           0.0154      0.1417      0.11      0.913
## Year2008          -0.0980      0.1303     -0.75      0.453
## Year2009          -0.0736      0.1380     -0.53      0.594
## Year2010          -0.1731      0.1372     -1.26      0.208
## Year2011          -0.1297      0.1331     -0.97      0.330
## Year2012           0.0197      0.1327      0.15      0.882
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0778
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 513 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.160  0.854  0.954  0.902  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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.366 1      1.538
## LastAuthorFemale  2.138 1      1.462
## Year              1.369 16      1.010

```

Residuals from first and last author



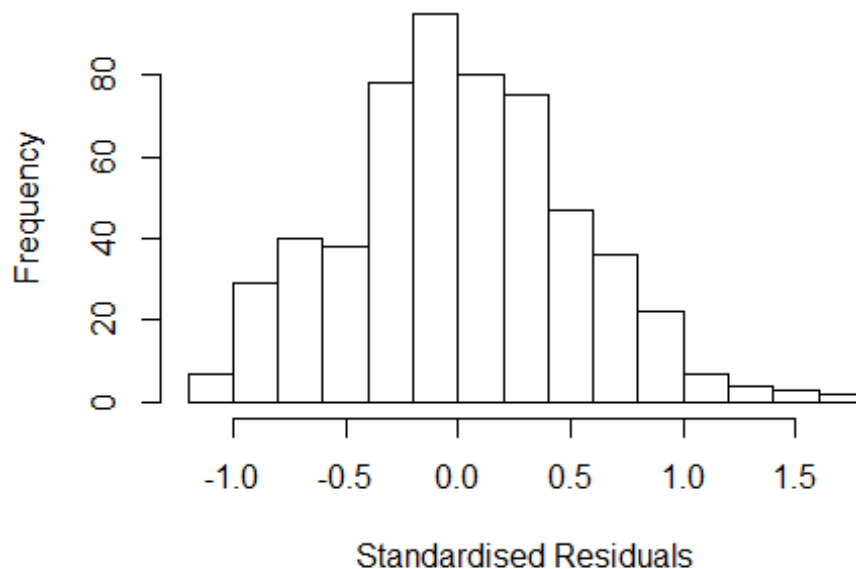
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17166 -0.31947 -0.00799  0.33439  1.68140
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.89456    0.10936   8.18 2e-15 ***
## FirstAuthorFemale1 0.15544    0.06736   2.31  0.021 *
## LastAuthorFemale1 0.05062    0.06633   0.76  0.446
## Year1997         0.01283    0.14748   0.09  0.931
## Year1998        -0.07666    0.13893  -0.55  0.581
## Year1999        -0.03257    0.15102  -0.22  0.829
## Year2000        -0.28130    0.13248  -2.12  0.034 *
## Year2001        -0.04907    0.14305  -0.34  0.732
## Year2002        -0.09896    0.13061  -0.76  0.449
## Year2003        -0.00171    0.14700  -0.01  0.991
## Year2004        -0.07333    0.13983  -0.52  0.600
## Year2005        -0.13968    0.18002  -0.78  0.438
```

```

## Year2006          0.17023      0.14758      1.15      0.249
## Year2007          0.05080      0.14265      0.36      0.722
## Year2008         -0.09259      0.13003     -0.71      0.477
## Year2009         -0.06340      0.13816     -0.46      0.646
## Year2010         -0.16009      0.14065     -1.14      0.256
## Year2011         -0.11963      0.13505     -0.89      0.376
## Year2012          0.07103      0.13504      0.53      0.599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.508
## Multiple R-squared:  0.0628, Adjusted R-squared:  0.0318
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.251  0.859   0.955   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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.244 1      1.115
## Year              1.244 16      1.007

```


Residuals from first author



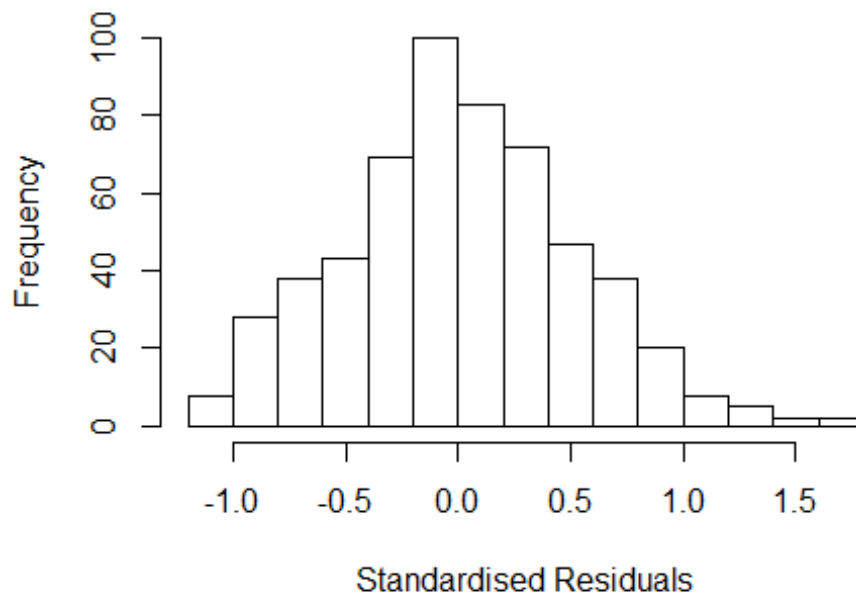
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1635 -0.3269 -0.0131 0.3420 1.6759
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89903 0.10864 8.28 9.9e-16 ***
## FirstAuthorFemale1 0.18915 0.04881 3.87 0.00012 ***
## Year1997 0.01462 0.14814 0.10 0.92141
## Year1998 -0.07573 0.13874 -0.55 0.58540
## Year1999 -0.02994 0.15012 -0.20 0.84201
## Year2000 -0.27714 0.13238 -2.09 0.03676 *
## Year2001 -0.05209 0.14342 -0.36 0.71657
## Year2002 -0.09797 0.13077 -0.75 0.45408
## Year2003 0.00655 0.14644 0.04 0.96435
## Year2004 -0.07231 0.14000 -0.52 0.60572
## Year2005 -0.14344 0.17934 -0.80 0.42417
## Year2006 0.16531 0.14633 1.13 0.25908
```

```

## Year2007          0.05810      0.14157      0.41  0.68166
## Year2008          -0.09063      0.12893     -0.70  0.48242
## Year2009          -0.06073      0.13883     -0.44  0.66196
## Year2010          -0.16355      0.13968     -1.17  0.24215
## Year2011          -0.11855      0.13524     -0.88  0.38110
## Year2012           0.07529      0.13455      0.56  0.57599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.0617, Adjusted R-squared:  0.0325
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.263  0.861  0.957  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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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          1.062
## Year              1.127 16          1.004

```

Residuals from last author



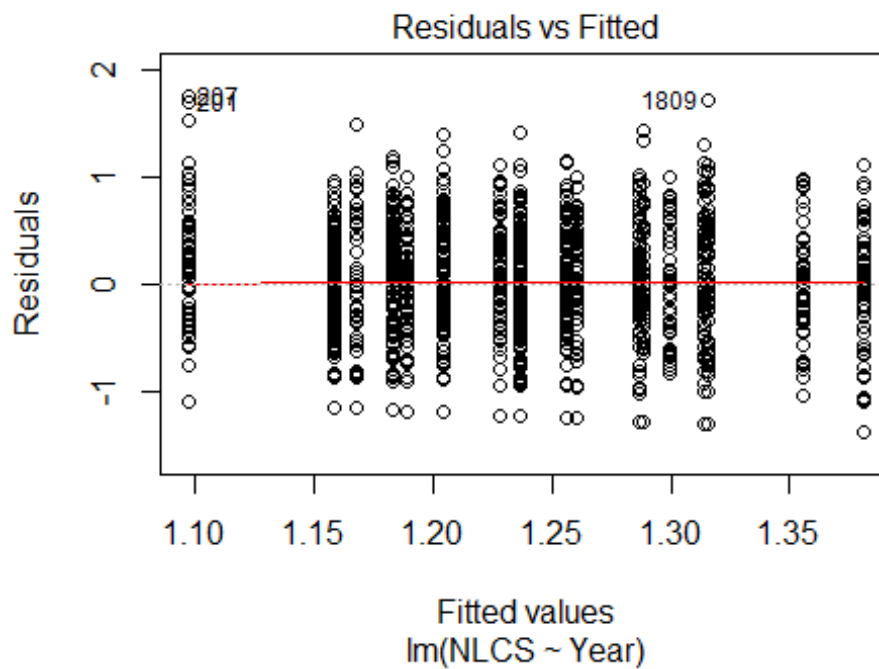
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15242 -0.34092 0.00883 0.34703 1.65705
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90608 0.11046 8.20 1.7e-15 ***
## LastAuthorFemale1 0.15549 0.04817 3.23 0.0013 **
## Year1997 0.00507 0.15017 0.03 0.9731
## Year1998 -0.07705 0.13832 -0.56 0.5777
## Year1999 -0.03967 0.15292 -0.26 0.7954
## Year2000 -0.27432 0.13273 -2.07 0.0392 *
## Year2001 -0.02335 0.14138 -0.17 0.8689
## Year2002 -0.08613 0.13030 -0.66 0.5089
## Year2003 -0.02224 0.14630 -0.15 0.8792
## Year2004 -0.05183 0.13963 -0.37 0.7106
## Year2005 -0.11847 0.18317 -0.65 0.5181
## Year2006 0.19086 0.15094 1.26 0.2066
```

```

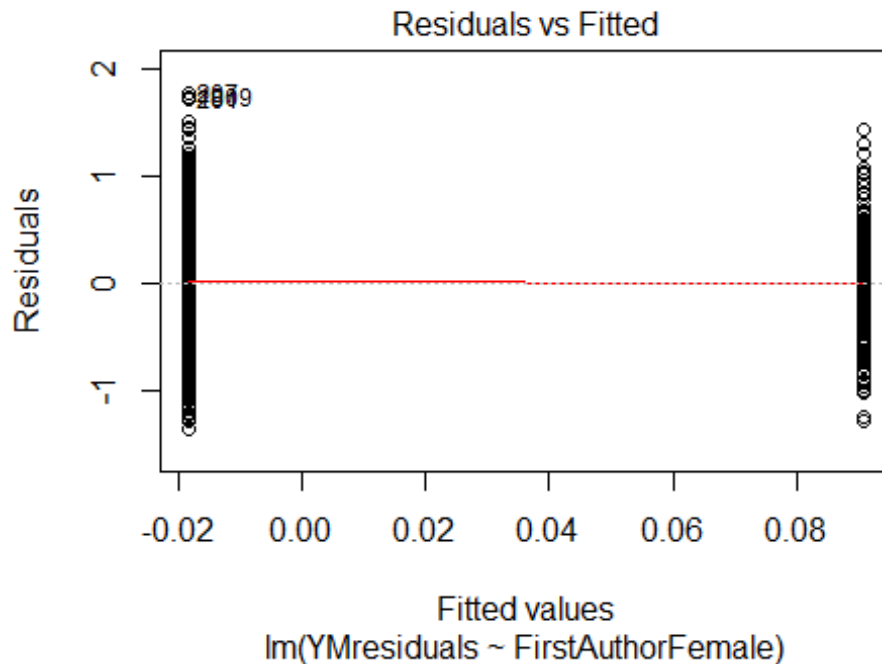
## Year2007      0.04244      0.14534      0.29      0.7704
## Year2008     -0.09492      0.13343     -0.71      0.4771
## Year2009     -0.04639      0.13725     -0.34      0.7355
## Year2010     -0.14221      0.14285     -1.00      0.3199
## Year2011     -0.10372      0.13508     -0.77      0.4429
## Year2012      0.09085      0.13715      0.66      0.5080
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.052, Adjusted R-squared:  0.0224
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 521 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.264  0.852  0.954  0.905  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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 563"
## [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
## 227 165 181 176 207 192 185 174 159 200 224 231 234 296 296
## 2011 2012
## 349 346
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 51 55 45 62 42 64 63 60 77 95 91 89 127 125
## 2011 2012

```

```
## 157 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 47 44 35 53 32 49 52 52 60 76 69 69 99 102
## 2011 2012
## 129 123
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 40, df = 16, p-value = 8e-04
```

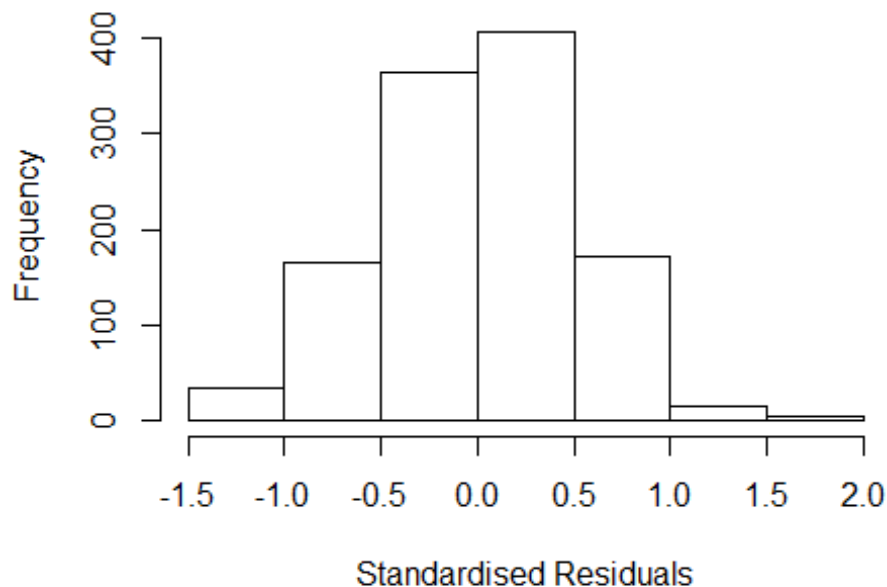


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.3, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 4.0223188543806"
## [1] "Male first author team size 2018 geometric mean: 3.46536666803656"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77435742986883"
## [1] "Male last author team size 2018 geometric mean: 3.57744217567246"
##
## 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 1.204 1      1.097
## LastAuthorFemale  1.109 1      1.053
## UniqueAuthors    1.491 4      1.051
## Year              1.616 16     1.015
```

Residuals from first and last author and team size



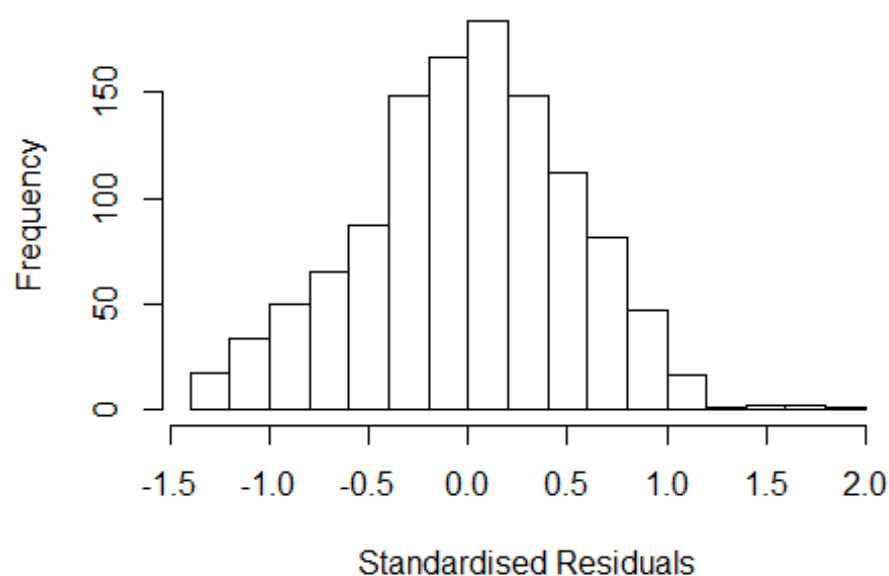
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4447 -0.3445 0.0179 0.3559 1.9489
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8481 0.0942 9.00 < 2e-16 ***
## FirstAuthorFemale1 0.0303 0.0401 0.75 0.450
## LastAuthorFemale1 0.0784 0.0476 1.65 0.100 .
## UniqueAuthors2 0.3216 0.0522 6.16 1.0e-09 ***
## UniqueAuthors3 0.2759 0.0534 5.17 2.8e-07 ***
## UniqueAuthors4 0.2759 0.0552 5.00 6.6e-07 ***
## UniqueAuthors5 0.3229 0.0583 5.53 3.9e-08 ***
## Year1997 0.2367 0.1111 2.13 0.033 *
## Year1998 0.2122 0.1088 1.95 0.051 .
## Year1999 0.0940 0.1373 0.68 0.494
```

```

## Year2000      0.2737      0.1111      2.46      0.014 *
## Year2001      0.3116      0.1359      2.29      0.022 *
## Year2002      0.1904      0.1276      1.49      0.136
## Year2003      0.2071      0.1253      1.65      0.099 .
## Year2004      0.1717      0.1134      1.51      0.130
## Year2005      0.0737      0.1023      0.72      0.471
## Year2006      0.2383      0.0993      2.40      0.017 *
## Year2007      0.1789      0.1058      1.69      0.091 .
## Year2008      0.1013      0.1046      0.97      0.333
## Year2009      0.1544      0.1022      1.51      0.131
## Year2010      0.0455      0.1001      0.45      0.649
## Year2011      0.0468      0.0951      0.49      0.623
## Year2012      0.0470      0.0934      0.50      0.615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.527
## Multiple R-squared:  0.075, Adjusted R-squared:  0.0571
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1052 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.142  0.873  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      8.61e-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 1.119 1      1.058
## LastAuthorFemale  1.102 1      1.050
## Year              1.169 16      1.005

```


Residuals from first and last author



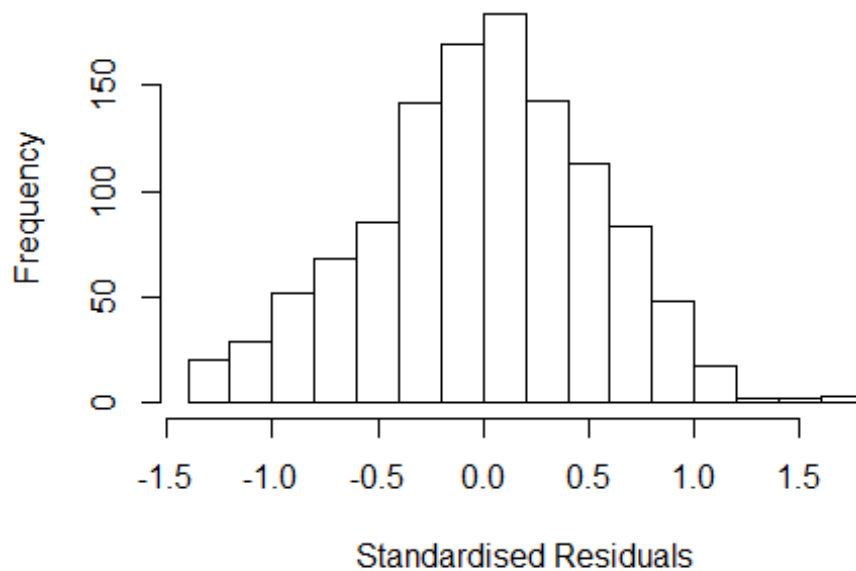
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3924 -0.3421  0.0102  0.3665  1.8041
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0582     0.0958   11.04  <2e-16 ***
## FirstAuthorFemale1  0.0708     0.0397    1.79   0.074 .
## LastAuthorFemale1  0.0943     0.0493    1.91   0.056 .
## Year1997          0.2245     0.1214    1.85   0.065 .
## Year1998          0.2214     0.1193    1.86   0.064 .
## Year1999          0.0510     0.1553    0.33   0.742
## Year2000          0.2867     0.1226    2.34   0.019 *
## Year2001          0.3436     0.1457    2.36   0.019 *
## Year2002          0.2396     0.1301    1.84   0.066 .
## Year2003          0.1747     0.1370    1.27   0.203
## Year2004          0.2229     0.1194    1.87   0.062 .
## Year2005          0.0854     0.1113    0.77   0.443
```

```

## Year2006          0.2227      0.1079      2.06      0.039 *
## Year2007          0.1881      0.1156      1.63      0.104
## Year2008          0.1400      0.1158      1.21      0.227
## Year2009          0.1691      0.1134      1.49      0.136
## Year2010          0.0766      0.1113      0.69      0.492
## Year2011          0.0843      0.1058      0.80      0.426
## Year2012          0.1069      0.1041      1.03      0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.026, Adjusted R-squared:  0.0107
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1059 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.220  0.861  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      8.61e-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 1.094 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
## -1.3431 -0.3457  0.0121  0.3603  1.7953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0587     0.0959   11.04  <2e-16 ***
## FirstAuthorFemale1  0.0817     0.0393    2.08   0.038 *
## Year1997          0.2320     0.1217    1.91   0.057 .
## Year1998          0.2262     0.1189    1.90   0.057 .
## Year1999          0.0522     0.1544    0.34   0.735
## Year2000          0.2844     0.1226    2.32   0.021 *
## Year2001          0.3470     0.1456    2.38   0.017 *
## Year2002          0.2471     0.1308    1.89   0.059 .
## Year2003          0.1889     0.1366    1.38   0.167
## Year2004          0.2307     0.1195    1.93   0.054 .
## Year2005          0.0917     0.1119    0.82   0.412
## Year2006          0.2299     0.1087    2.11   0.035 *
```

```

## Year2007          0.1949      0.1160      1.68      0.093 .
## Year2008          0.1472      0.1158      1.27      0.204
## Year2009          0.1790      0.1137      1.57      0.116
## Year2010          0.0872      0.1113      0.78      0.434
## Year2011          0.0982      0.1053      0.93      0.351
## Year2012          0.1195      0.1037      1.15      0.250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0231, Adjusted R-squared:  0.00856
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1061 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.226  0.862  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      8.61e-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 1.076 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 ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.3564 -0.3398  0.0106  0.3572  1.7856

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0684    0.0960   11.13 <2e-16 ***
## LastAuthorFemale1 0.1059    0.0487    2.18  0.030 *
## Year1997          0.2245    0.1217    1.84  0.065 .
## Year1998          0.2181    0.1196    1.82  0.068 .
## Year1999          0.0522    0.1577    0.33  0.741
## Year2000          0.2880    0.1229    2.34  0.019 *
## Year2001          0.3339    0.1463    2.28  0.023 *
## Year2002          0.2363    0.1307    1.81  0.071 .
## Year2003          0.1859    0.1373    1.35  0.176
## Year2004          0.2213    0.1200    1.84  0.065 .
## Year2005          0.0873    0.1123    0.78  0.437
## Year2006          0.2211    0.1085    2.04  0.042 *
## Year2007          0.1850    0.1158    1.60  0.111
## Year2008          0.1441    0.1160    1.24  0.214
## Year2009          0.1716    0.1139    1.51  0.132
## Year2010          0.0755    0.1119    0.67  0.500
## Year2011          0.0931    0.1063    0.88  0.382
## Year2012          0.1064    0.1045    1.02  0.309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.00922
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.230  0.862  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      8.61e-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: 1161"
## [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
## 15 16 12 15 13 17 7 8 12 9 20 17 14 7 19
## 2011 2012
## 10 7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 4 3 6 2 1 1 2 4 7 4 7 8 4 9
## 2011 2012
## 6 4
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 4 3 6 2 1 1 0 2 4 3 5 7 1 7
## 2011 2012
## 3 3
## [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: 3.57649722014261"

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

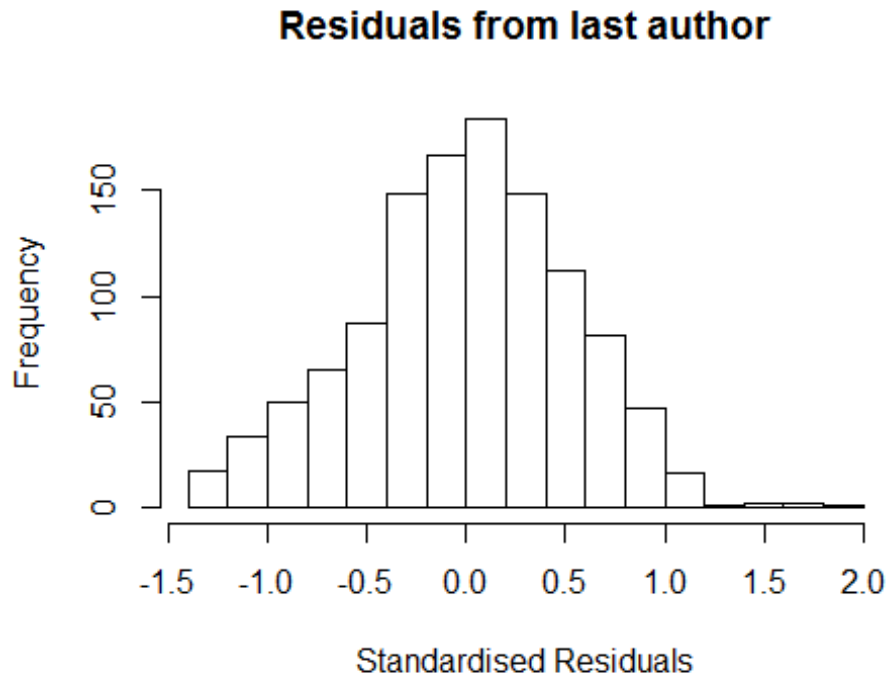
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.46410161513775"
## [1] "Male last author team size 2018 geometric mean: 3.55897113644341"

## 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 = 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"

```

```
## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'LastAuthorFemale1'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```



```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -4.481e+15  1      NaN
## LastAuthorFemale -9.638e+15  1      NaN
## UniqueAuthors    -2.929e+04  4      NaN
## Year              8.071e+32 15     12.5

## [1] "List of 0 outliers with 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
## -2.39e+00 -2.30e-01 -5.55e-16  1.90e-01  9.31e-01
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept)      0.8236      0.5926      1.39      0.1739
## FirstAuthorFemale1 -0.0927      0.1583     -0.59      0.5622
## LastAuthorFemale1  2.2033      0.2871      7.67      7.8e-09 ***
## UniqueAuthors2      0.1939      0.2042      0.95      0.3493
## UniqueAuthors3      0.1388      0.2303      0.60      0.5508
## UniqueAuthors4      0.7568      0.2591      2.92      0.0062 **
## UniqueAuthors5      0.7174      0.2531      2.83      0.0078 **
## Year1997            0.4838      0.6014      0.80      0.4269
## Year1998            0.6577      0.6198      1.06      0.2963
## Year1999            0.5513      0.6010      0.92      0.3656
## Year2000            0.6524      0.4711      1.38      0.1754
## Year2001           -0.5107      0.5370     -0.95      0.3485
## Year2002            0.1306      0.5252      0.25      0.8052
## Year2004           -0.4458      0.5549     -0.80      0.4275
## Year2005            0.3337      0.5359      0.62      0.5378
## Year2006           -0.0411      0.5981     -0.07      0.9457
## Year2007           -0.1297      0.5083     -0.26      0.8002
## Year2008            0.0100      0.5250      0.02      0.9849
## Year2009            0.3566      0.5252      0.68      0.5019
## Year2010           -0.4029      0.5416     -0.74      0.4622
## Year2011           -0.3936      0.4944     -0.80      0.4316
## Year2012            0.5464      0.6268      0.87      0.3897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.658, Adjusted R-squared:  0.44
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 37 is an outlier with |weight| = 0 ( < 0.0018);
## 12 weights are ~= 1. The remaining 42 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.684  0.918  0.954  0.937  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)

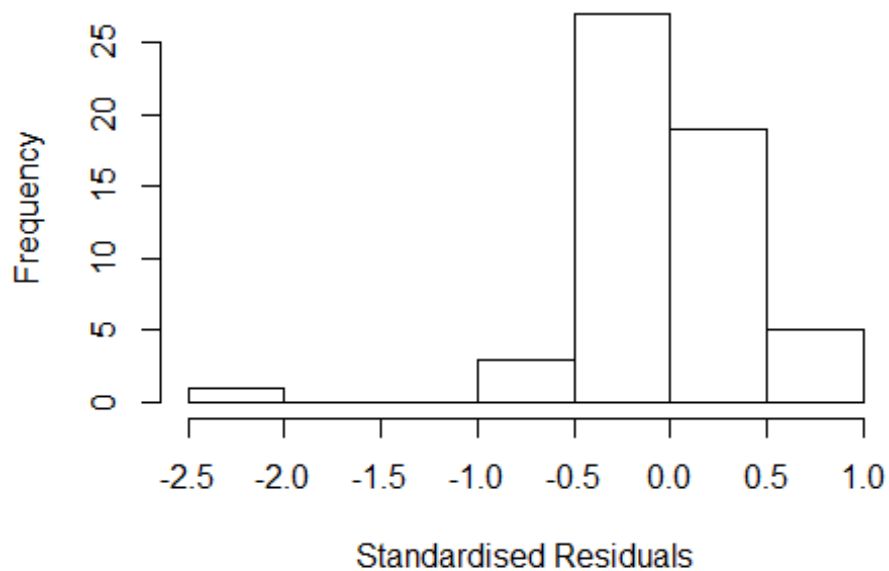
```



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

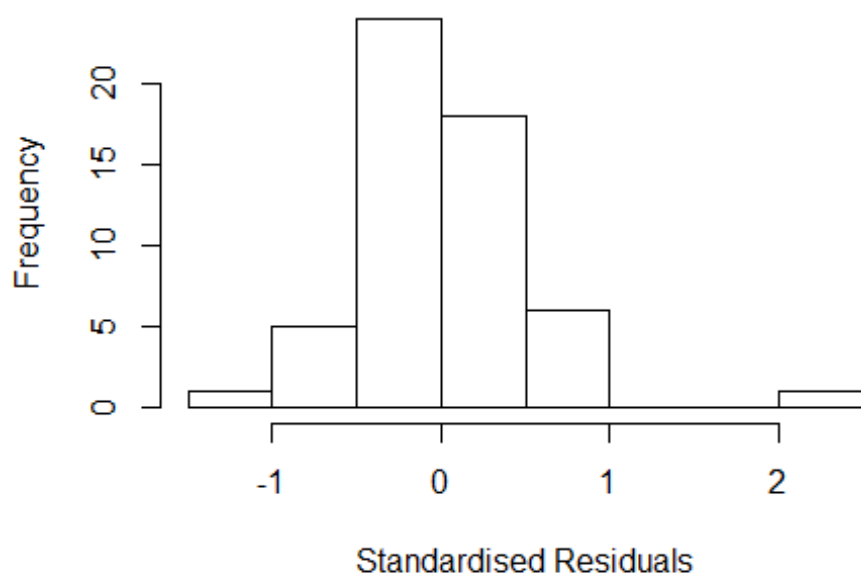
## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'LastAuthorFemale1'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```

Residuals from first and last author and team size



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.476e+12 1      1573435
## LastAuthorFemale -5.346e+14 1           NaN
## Year              -1.302e+28 15           NaN
```

Residuals from first and last author



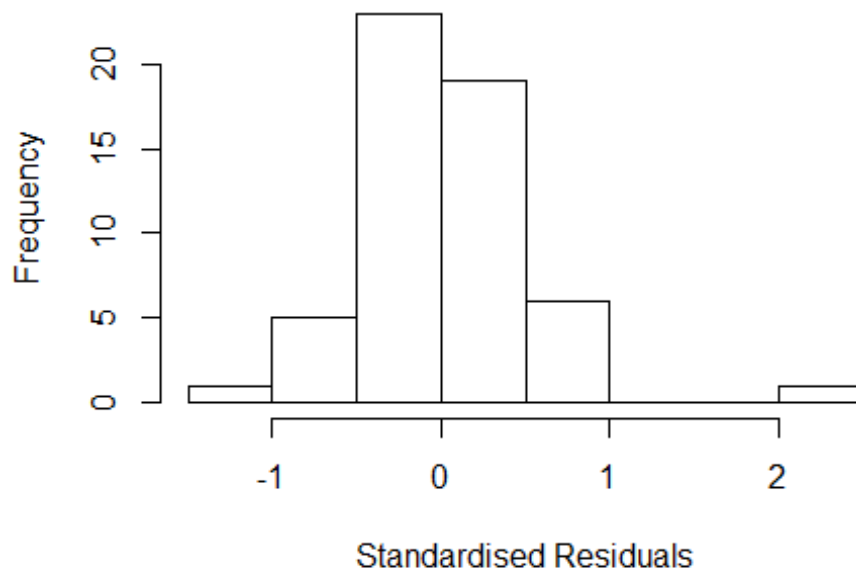
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27e+00 -2.53e-01 4.58e-16 2.67e-01 2.21e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.271 0.580 2.19 0.035 *
## FirstAuthorFemale1 -0.109 0.172 -0.63 0.530
## LastAuthorFemale1 0.200 0.187 1.07 0.292
## Year1997 0.328 0.614 0.53 0.596
## Year1998 0.309 0.631 0.49 0.627
## Year1999 0.214 0.629 0.34 0.736
## Year2000 0.689 0.618 1.11 0.272
## Year2001 -0.747 0.590 -1.27 0.213
## Year2002 -0.123 0.580 -0.21 0.834
## Year2004 -0.429 0.764 -0.56 0.578
## Year2005 0.347 0.587 0.59 0.559
## Year2006 -0.488 0.593 -0.82 0.416
```

```

## Year2007          -0.170      0.582   -0.29    0.772
## Year2008          -0.102      0.613   -0.17    0.869
## Year2009           0.103      0.580    0.18    0.860
## Year2010          -0.636      0.673   -0.94    0.351
## Year2011          -0.477      0.592   -0.81    0.425
## Year2012           0.373      0.848    0.44    0.662
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.477
## Multiple R-squared:  0.434, Adjusted R-squared:  0.174
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 28 is an outlier with |weight| <= 0.00032 ( < 0.0018);
## 11 weights are ~= 1. The remaining 43 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.458  0.900  0.966  0.909  0.982  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      1.82e-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 1.683e+14 1      1.297e+07
## Year              1.683e+14 15      2.980e+00

```

Residuals from first author



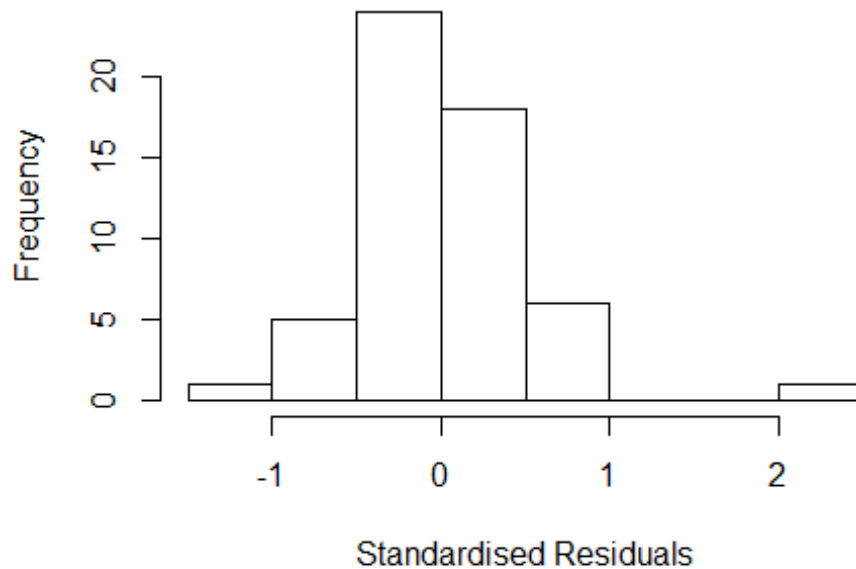
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29e+00 -2.64e-01 -3.61e-16 2.64e-01 2.41e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2854 0.5730 2.24 0.031 *
## FirstAuthorFemale1 -0.1105 0.1663 -0.66 0.510
## Year1997 0.3174 0.6064 0.52 0.604
## Year1998 0.2944 0.6261 0.47 0.641
## Year1999 0.1982 0.6236 0.32 0.752
## Year2000 0.6748 0.6123 1.10 0.277
## Year2001 -0.7609 0.5839 -1.30 0.200
## Year2002 -0.1374 0.5730 -0.24 0.812
## Year2004 -0.4432 0.7651 -0.58 0.566
## Year2005 0.3322 0.5808 0.57 0.571
## Year2006 -0.5029 0.5786 -0.87 0.390
## Year2007 -0.1845 0.5751 -0.32 0.750
```

```

## Year2008          -0.0828      0.5920   -0.14    0.889
## Year2009           0.0886      0.5730    0.15    0.878
## Year2010          -0.6543      0.6675   -0.98    0.333
## Year2011          -0.4919      0.5847   -0.84    0.405
## Year2012           0.3595      0.9246    0.39    0.700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.436, Adjusted R-squared:  0.199
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 28 is an outlier with |weight| = 0 ( < 0.0018);
## 9 weights are ~= 1. The remaining 45 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.431  0.896  0.966  0.909  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      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 1.007 1          1.003
## Year            1.007 15          1.000

```

Residuals from last author



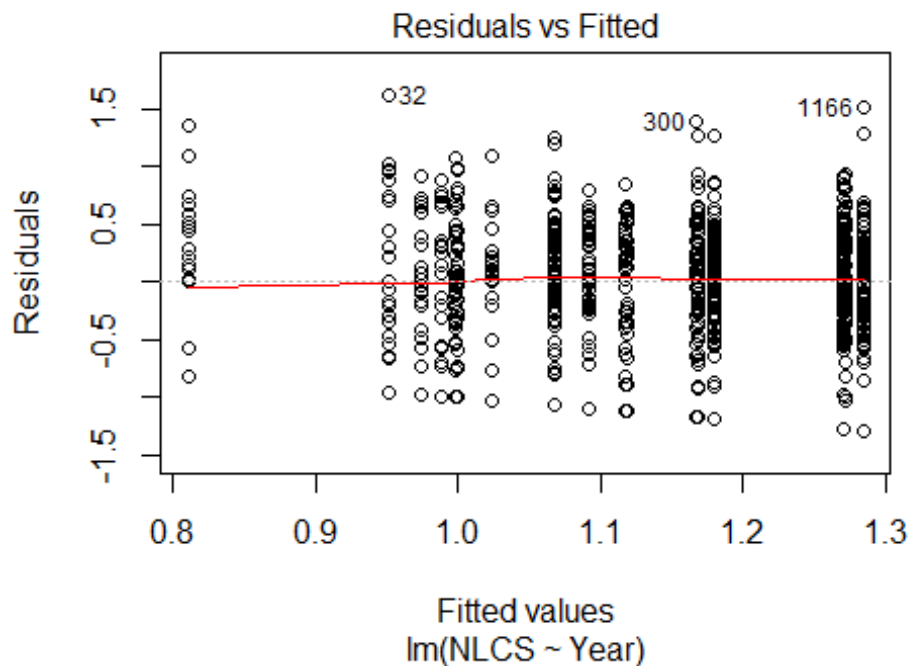
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25e+00 -2.97e-01 6.94e-17 2.78e-01 9.57e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.249 0.585 2.14 0.039 *
## LastAuthorFemale1 1.149 1.425 0.81 0.425
## Year1997 0.310 0.612 0.51 0.615
## Year1998 0.331 0.631 0.52 0.604
## Year1999 0.217 0.639 0.34 0.736
## Year2000 0.656 0.629 1.04 0.304
## Year2001 -0.835 0.585 -1.43 0.161
## Year2002 -0.101 0.585 -0.17 0.863
## Year2004 -0.462 0.720 -0.64 0.525
## Year2005 0.339 0.598 0.57 0.574
## Year2006 -0.136 0.700 -0.19 0.847
## Year2007 -0.169 0.589 -0.29 0.775
```

```

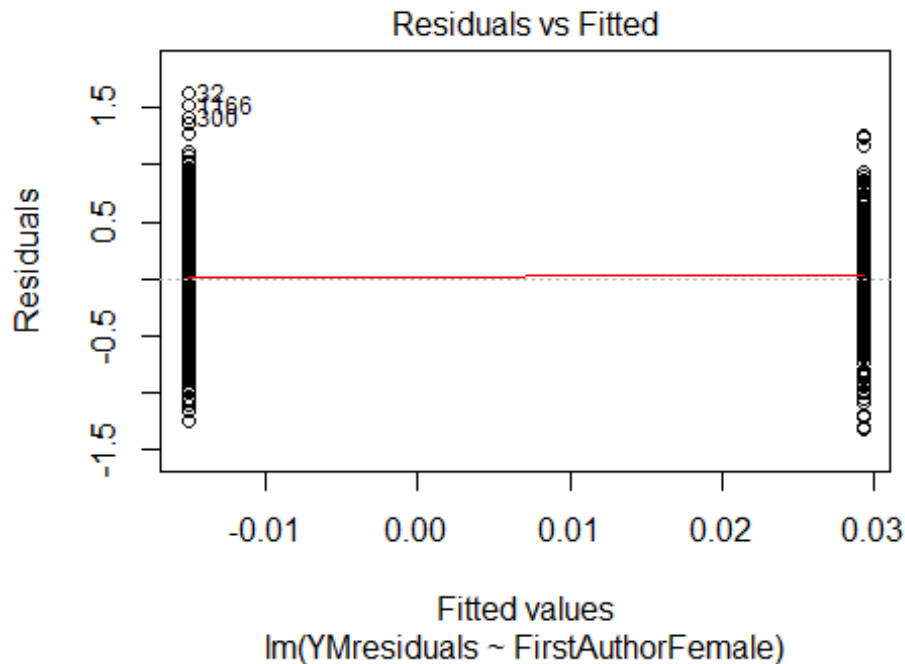
## Year2008          -0.207          0.621      -0.33      0.741
## Year2009           0.125          0.585       0.21      0.832
## Year2010          -0.608          0.674      -0.90      0.372
## Year2011          -0.456          0.596      -0.76      0.449
## Year2012           0.394          0.867       0.45      0.652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.429, Adjusted R-squared:  0.189
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 47 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.497  0.914  0.957  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      1.82e-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 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
##   53   54   45   35   48   39   50   43   56   58   72   93  128  162  187
## 2011 2012
##  159  174
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   25   20   24   19   21   19   26   24   26   32   45   39   66   82   93
## 2011 2012
##   90   95

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   18   21   15   17   17   25   23   24   28   41   31   56   71   83
## 2011 2012
##   78   89
## [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
```



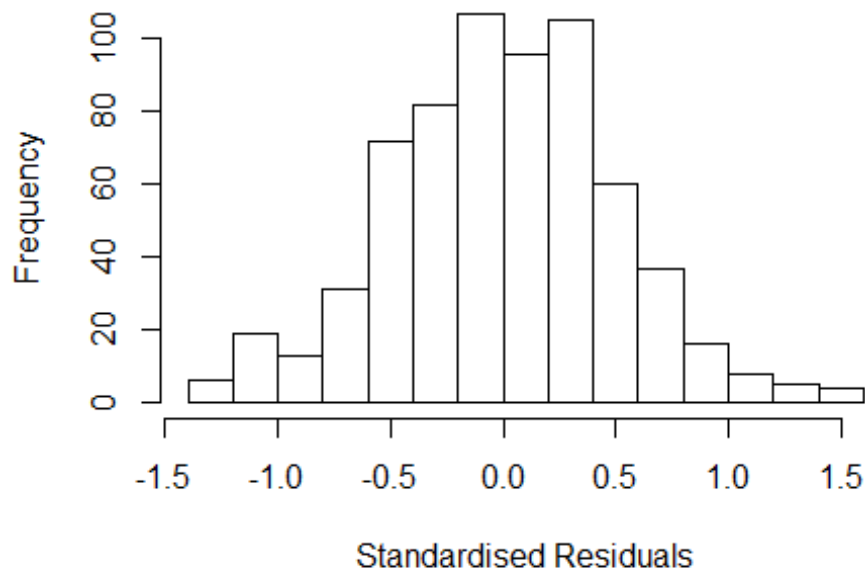
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.29, df = 1, p-value = 0.6
## [1] "Female first author team size 2018 geometric mean: 5.59857729137298"
## [1] "Male first author team size 2018 geometric mean: 4.68425747110017"
## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.99168792105753"
## [1] "Male last author team size 2018 geometric mean: 5.15848657407405"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 230, 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	1.158	1	1.076
LastAuthorFemale	1.188	1	1.090
UniqueAuthors	2.406	4	1.116
Year	2.607	16	1.030

Residuals from first and last author and team size



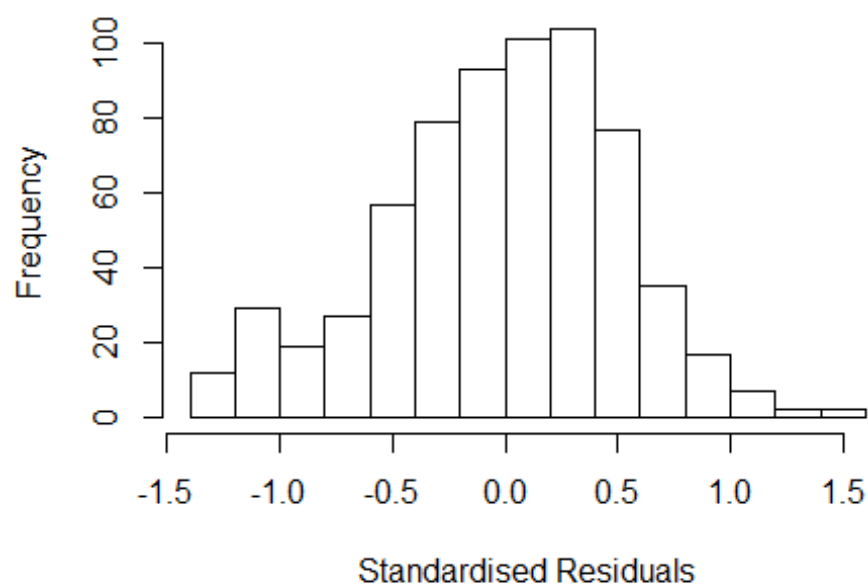
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26747 -0.33329 0.00186 0.32384 1.59733
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3677 0.2337 1.57 0.12
## FirstAuthorFemale1 0.0173 0.0446 0.39 0.70
## LastAuthorFemale1 0.0708 0.0562 1.26 0.21
## UniqueAuthors2 0.5975 0.1257 4.75 2.5e-06 ***
## UniqueAuthors3 0.6600 0.1107 5.96 4.1e-09 ***
## UniqueAuthors4 0.7233 0.1125 6.43 2.5e-10 ***
## UniqueAuthors5 0.7071 0.1117 6.33 4.7e-10 ***
## Year1997 -0.0318 0.2693 -0.12 0.91
## Year1998 0.1650 0.2492 0.66 0.51
## Year1999 -0.0286 0.2389 -0.12 0.90
```

```

## Year2000          0.0337      0.2756      0.12      0.90
## Year2001          0.2225      0.2480      0.90      0.37
## Year2002         -0.0798      0.2661     -0.30      0.76
## Year2003          0.1644      0.2326      0.71      0.48
## Year2004          0.2141      0.2411      0.89      0.37
## Year2005         -0.0299      0.2346     -0.13      0.90
## Year2006          0.0632      0.2286      0.28      0.78
## Year2007          0.2364      0.2309      1.02      0.31
## Year2008          0.2083      0.2248      0.93      0.35
## Year2009          0.2402      0.2206      1.09      0.28
## Year2010          0.1173      0.2209      0.53      0.60
## Year2011          0.2460      0.2235      1.10      0.27
## Year2012          0.0603      0.2215      0.27      0.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.198, Adjusted R-squared:  0.17
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 606 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.205  0.861  0.946  0.893  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      1.51e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1      1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.126 1      1.061
## LastAuthorFemale  1.124 1      1.060
## Year              1.198 16      1.006

```

Residuals from first and last author



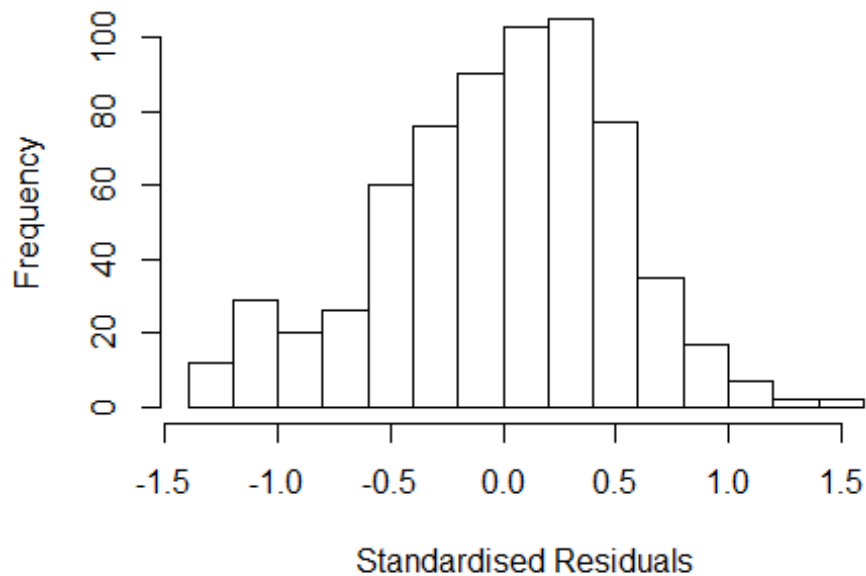
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3564 -0.3372 0.0316 0.3446 1.5259
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8197 0.1767 4.64 4.2e-06 ***
## FirstAuthorFemale1 0.0462 0.0457 1.01 0.313
## LastAuthorFemale1 0.0361 0.0611 0.59 0.555
## Year1997 0.0939 0.2416 0.39 0.698
## Year1998 0.2057 0.2448 0.84 0.401
## Year1999 0.0921 0.2232 0.41 0.680
## Year2000 0.2040 0.2193 0.93 0.353
## Year2001 0.3810 0.2224 1.71 0.087 .
## Year2002 -0.0729 0.2391 -0.30 0.761
## Year2003 0.3366 0.2090 1.61 0.108
## Year2004 0.3549 0.2024 1.75 0.080 .
## Year2005 0.1379 0.2039 0.68 0.499
```

```

## Year2006          0.2864      0.1938      1.48      0.140
## Year2007          0.4350      0.1995      2.18      0.030 *
## Year2008          0.4184      0.1890      2.21      0.027 *
## Year2009          0.4544      0.1830      2.48      0.013 *
## Year2010          0.3215      0.1832      1.75      0.080 .
## Year2011          0.4391      0.1886      2.33      0.020 *
## Year2012          0.2537      0.1893      1.34      0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0696, Adjusted R-squared:  0.0436
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 605 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.326  0.872  0.949  0.897  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      1.51e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.095 1      1.046
## Year              1.095 16      1.003

```

Residuals from first author



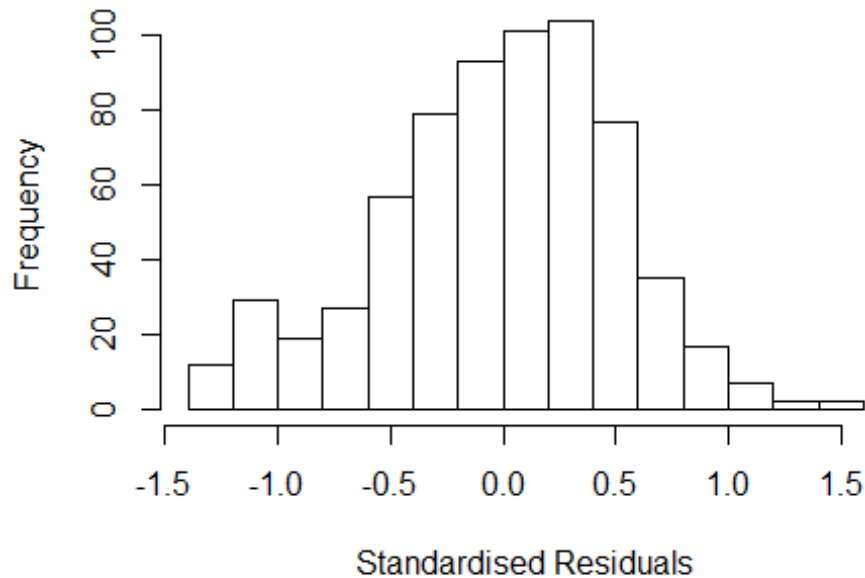
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3304 -0.3326 0.0319 0.3467 1.5220
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8263 0.1760 4.70 3.3e-06 ***
## FirstAuthorFemale1 0.0524 0.0453 1.16 0.247
## Year1997 0.0911 0.2401 0.38 0.705
## Year1998 0.2016 0.2450 0.82 0.411
## Year1999 0.0900 0.2233 0.40 0.687
## Year2000 0.1979 0.2197 0.90 0.368
## Year2001 0.3766 0.2224 1.69 0.091 .
## Year2002 -0.0721 0.2402 -0.30 0.764
## Year2003 0.3284 0.2084 1.58 0.116
## Year2004 0.3511 0.2023 1.74 0.083 .
## Year2005 0.1374 0.2038 0.67 0.501
## Year2006 0.2840 0.1936 1.47 0.143
```

```

## Year2007          0.4292      0.1992      2.15      0.032 *
## Year2008          0.4182      0.1891      2.21      0.027 *
## Year2009          0.4516      0.1828      2.47      0.014 *
## Year2010          0.3188      0.1831      1.74      0.082 .
## Year2011          0.4366      0.1885      2.32      0.021 *
## Year2012          0.2483      0.1885      1.32      0.188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.0688, Adjusted R-squared:  0.0442
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 606 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.331  0.874  0.949  0.898  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      1.51e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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
## -1.3340 -0.3400 0.0284 0.3401 1.5154
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8294 0.1764 4.70 3.1e-06 ***
## LastAuthorFemale1 0.0494 0.0604 0.82 0.414
## Year1997 0.0926 0.2427 0.38 0.703
## Year1998 0.2137 0.2453 0.87 0.384
## Year1999 0.0903 0.2233 0.40 0.686
## Year2000 0.2160 0.2173 0.99 0.320
## Year2001 0.3832 0.2230 1.72 0.086 .
## Year2002 -0.0653 0.2383 -0.27 0.784
## Year2003 0.3508 0.2071 1.69 0.091 .
## Year2004 0.3560 0.2028 1.76 0.080 .
## Year2005 0.1396 0.2028 0.69 0.491
## Year2006 0.2905 0.1933 1.50 0.133
```

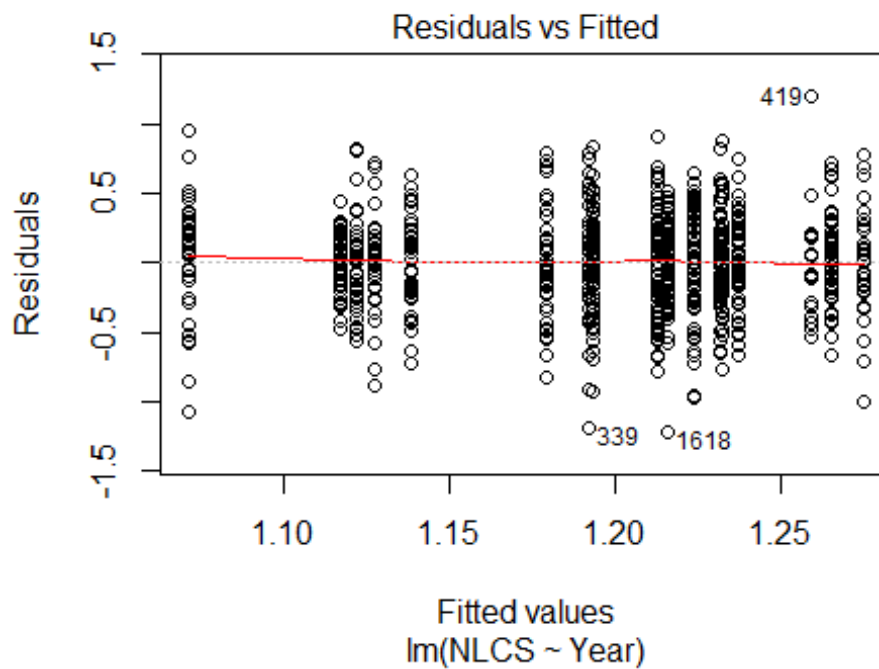


```

## Year2007          0.4418      0.1992      2.22      0.027 *
## Year2008          0.4226      0.1883      2.24      0.025 *
## Year2009          0.4552      0.1827      2.49      0.013 *
## Year2010          0.3260      0.1826      1.79      0.075 .
## Year2011          0.4439      0.1878      2.36      0.018 *
## Year2012          0.2577      0.1886      1.37      0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0677, Adjusted R-squared:  0.0431
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 605 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.334  0.874   0.949   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      1.51e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 661"
## [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
##   67   73   74   58   63   55   47   65   58   55   61   63   60   95  107
## 2011 2012
##  121  114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   50   56   55   39   35   18   35   42   38   42   43   47   43   59   71
## 2011 2012

```

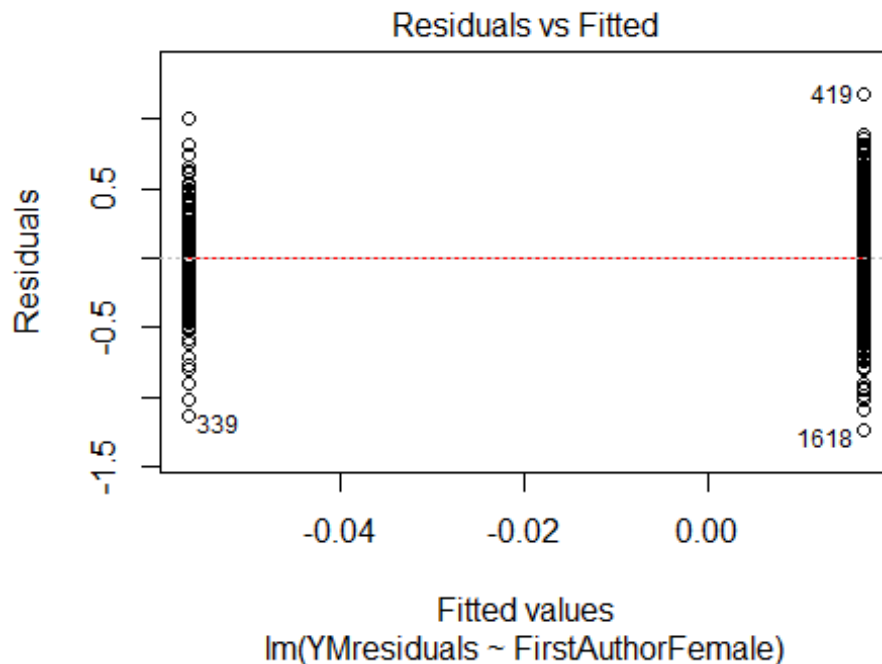
```
## 78 67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 48 50 51 38 30 14 34 38 31 34 39 41 37 53 66
## 2011 2012
## 68 56
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.008
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.91, df = 1, p-value = 0.3
## [1] "Female first author team size 2018 geometric mean: 4.3856828944377"
## [1] "Male first author team size 2018 geometric mean: 4.0599981710681"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

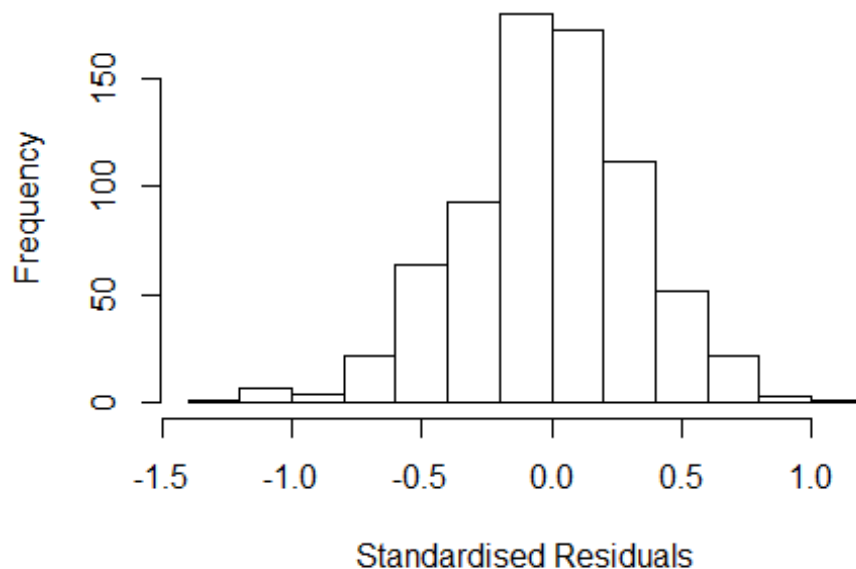
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 460, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.59571601874489"
## [1] "Male last author team size 2018 geometric mean: 4.32028517768251"

## 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 = 270, 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 1.154 1      1.074
## LastAuthorFemale  1.127 1      1.062
## UniqueAuthors    1.734 4      1.071
## Year              1.947 16     1.021
```

Residuals from first and last author and team size



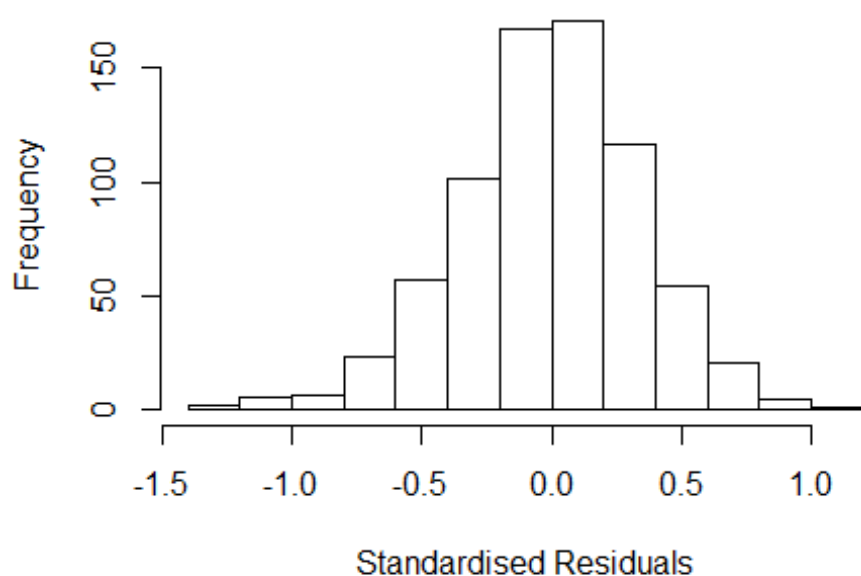
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29628 -0.21847 -0.00488  0.21640  1.06614
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23703    0.07650   16.17  <2e-16 ***
## FirstAuthorFemale1 -0.09872    0.03377   -2.92  0.0036 **
## LastAuthorFemale1 -0.06629    0.05086   -1.30  0.1929
## UniqueAuthors2    -0.08769    0.07233   -1.21  0.2258
## UniqueAuthors3     0.01376    0.06795    0.20  0.8396
## UniqueAuthors4    -0.01974    0.06835   -0.29  0.7728
## UniqueAuthors5     0.06928    0.06850    1.01  0.3122
## Year1997           0.05709    0.06536    0.87  0.3827
## Year1998           0.04244    0.06058    0.70  0.4837
## Year1999          -0.09345    0.06840   -1.37  0.1723
```

```

## Year2000      0.05042    0.08802    0.57    0.5670
## Year2001      0.08255    0.13955    0.59    0.5543
## Year2002      0.04942    0.08033    0.62    0.5386
## Year2003     -0.03098    0.07178   -0.43    0.6662
## Year2004     -0.05277    0.07777   -0.68    0.4977
## Year2005     -0.01784    0.08446   -0.21    0.8328
## Year2006      0.02333    0.06454    0.36    0.7178
## Year2007     -0.09172    0.05680   -1.61    0.1068
## Year2008     -0.05724    0.07842   -0.73    0.4657
## Year2009      0.04559    0.06913    0.66    0.5097
## Year2010     -0.00557    0.06846   -0.08    0.9351
## Year2011     -0.02891    0.06246   -0.46    0.6436
## Year2012      0.05627    0.05935    0.95    0.3435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.322
## Multiple R-squared:  0.0621, Adjusted R-squared:  0.0329
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 662 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0692 0.8580 0.9480 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          1.37e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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 1.072 1 1.036
## Year 1.199 16 1.006

```

Residuals from first and last author



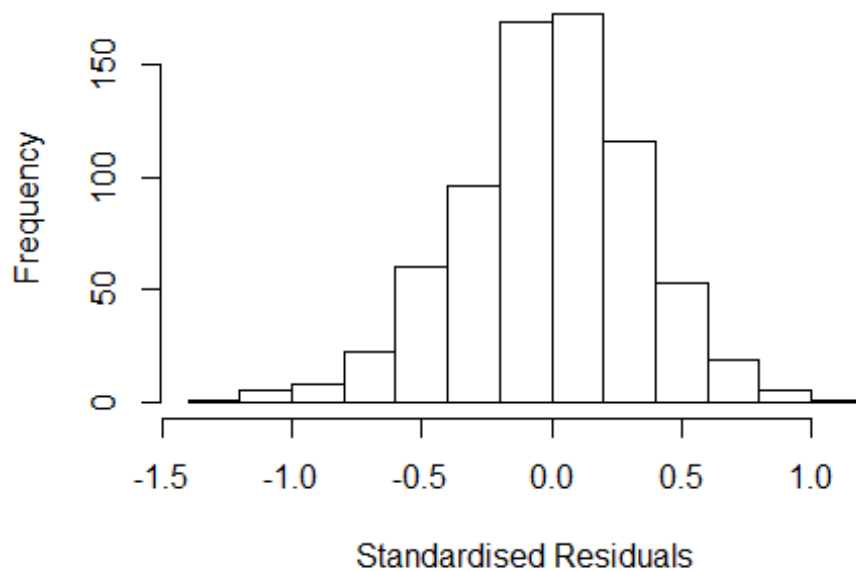
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22291 -0.21484  0.00108  0.22364  1.16859
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.22838    0.04638   26.48  <2e-16 ***
## FirstAuthorFemale1 -0.08770    0.03333   -2.63   0.0087 **
## LastAuthorFemale1  -0.06382    0.05139   -1.24   0.2147
## Year1997           0.06324    0.06563    0.96   0.3356
## Year1998           0.03912    0.06250    0.63   0.5316
## Year1999          -0.09486    0.06992   -1.36   0.1753
## Year2000           0.04658    0.08728    0.53   0.5938
## Year2001           0.05803    0.12740    0.46   0.6489
## Year2002           0.08128    0.07637    1.06   0.2875
## Year2003          -0.02854    0.07289   -0.39   0.6955
## Year2004          -0.03738    0.07633   -0.49   0.6245
## Year2005          -0.02538    0.08364   -0.30   0.7616
```

```

## Year2006          0.03895    0.06685    0.58    0.5603
## Year2007          -0.08042    0.05787   -1.39    0.1651
## Year2008          -0.04719    0.07886   -0.60    0.5497
## Year2009           0.06978    0.06972    1.00    0.3172
## Year2010           0.01515    0.07020    0.22    0.8292
## Year2011          -0.00162    0.06322   -0.03    0.9796
## Year2012           0.05834    0.06188    0.94    0.3460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared:  0.0374, Adjusted R-squared:  0.013
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 663 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.124  0.859   0.951   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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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      1.004

```

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
## -1.27781 -0.21268 0.00348 0.21938 1.17696
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22533 0.04652 26.34 <2e-16 ***
## FirstAuthorFemale1 -0.08974 0.03338 -2.69 0.0073 **
## Year1997 0.06173 0.06638 0.93 0.3527
## Year1998 0.03815 0.06286 0.61 0.5441
## Year1999 -0.09559 0.07018 -1.36 0.1736
## Year2000 0.04541 0.08805 0.52 0.6062
## Year2001 0.05272 0.12594 0.42 0.6756
## Year2002 0.07869 0.07632 1.03 0.3029
## Year2003 -0.02882 0.07252 -0.40 0.6912
## Year2004 -0.04207 0.07674 -0.55 0.5838
## Year2005 -0.02841 0.08308 -0.34 0.7325
## Year2006 0.03824 0.06713 0.57 0.5691
```



```

## Year2007          -0.08149    0.05819   -1.40    0.1618
## Year2008          -0.05022    0.07914   -0.63    0.5259
## Year2009           0.07047    0.07034    1.00    0.3168
## Year2010           0.01202    0.07072    0.17    0.8650
## Year2011          -0.00568    0.06317   -0.09    0.9284
## Year2012           0.05248    0.06196    0.85    0.3973
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared:  0.0349, Adjusted R-squared:  0.0118
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 661 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0851 0.8600 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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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

## [1] "List of 0 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
## -1.2511 -0.2148  0.0072  0.2070  1.1950

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21656    0.04636   26.24  <2e-16 ***
## LastAuthorFemale1 -0.07003    0.05207   -1.35    0.18
## Year1997        0.06193    0.06743    0.92    0.36
## Year1998        0.03532    0.06345    0.56    0.58
## Year1999       -0.10411    0.07112   -1.46    0.14
## Year2000        0.03459    0.08406    0.41    0.68
## Year2001        0.04343    0.12029    0.36    0.72
## Year2002        0.07247    0.07676    0.94    0.35
## Year2003       -0.04146    0.07445   -0.56    0.58
## Year2004       -0.04440    0.07819   -0.57    0.57
## Year2005       -0.01758    0.08217   -0.21    0.83
## Year2006        0.03742    0.06716    0.56    0.58
## Year2007       -0.08825    0.05853   -1.51    0.13
## Year2008       -0.05969    0.07652   -0.78    0.44
## Year2009        0.06295    0.06864    0.92    0.36
## Year2010        0.00686    0.07047    0.10    0.92
## Year2011       -0.00863    0.06310   -0.14    0.89
## Year2012        0.04322    0.06194    0.70    0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.327
## Multiple R-squared:  0.0256, Adjusted R-squared:  0.00229
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 660 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.111  0.861  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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 728"
## [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 2010 2011
##    2    3    3    2    7    5    5    6    6    5   11   11    8    2    1
## 2012
##    2
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2010 2011
##    1    2    2    1    5    1    1    3    6    2    7    7    3    1    0
## 2012
##    2
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2010 2011
##    1    2    2    1    5    1    1    2    6    2    7    7    2    1    0
## 2012
##    2
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5.31329284591305"
## [1] "Male first author team size 2018 geometric mean: 4.47213595499958"

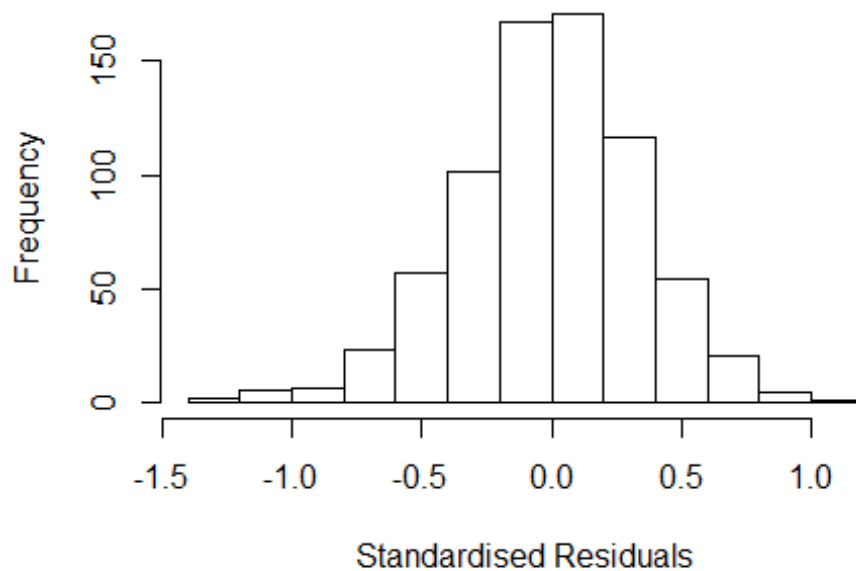
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5"
## [1] "Male last author team size 2018 geometric mean: 4.94923200383977"

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

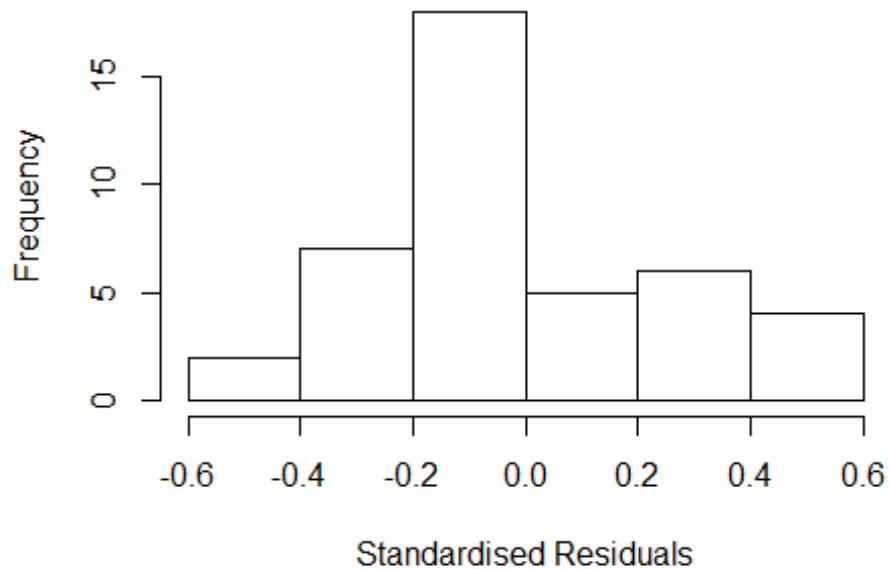
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, 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 1.954e+00 1          1.398
## LastAuthorFemale -5.131e+14 1          NaN
## UniqueAuthors    -2.379e+31 4          NaN
## Year             -6.973e+45 14         NaN
```

Residuals from first and last author and team size



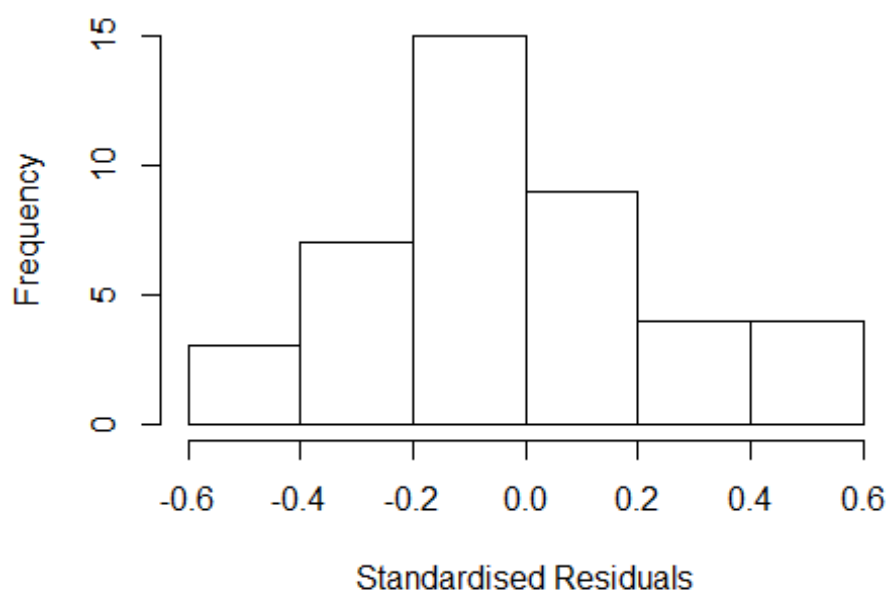
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.49842 -0.12043 -0.00165 0.16635 0.59332
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1459 0.1816 6.31 2.9e-06 ***
## FirstAuthorFemale1 0.0785 0.1487 0.53 0.60293
## LastAuthorFemale1 0.2077 0.1214 1.71 0.10197
## UniqueAuthors2 -0.0605 0.1654 -0.37 0.71813
## UniqueAuthors3 -0.0961 0.1829 -0.53 0.60472
## UniqueAuthors4 -0.0885 0.2271 -0.39 0.70078
## UniqueAuthors5 -0.5287 0.2965 -1.78 0.08899 .
## Year1997 0.6474 0.2304 2.81 0.01049 *
## Year1998 0.7882 0.4301 1.83 0.08108 .
## Year1999 0.2701 0.1816 1.49 0.15164
```

```

## Year2000          0.0591      0.1443      0.41  0.68628
## Year2001          1.1373      0.2069      5.50  1.9e-05 ***
## Year2002          0.2457      0.1214      2.02  0.05598 .
## Year2003          0.1296      0.2489      0.52  0.60799
## Year2004          0.3773      0.1712      2.20  0.03885 *
## Year2005          1.0189      0.1607      6.34  2.8e-06 ***
## Year2006          0.2620      0.1530      1.71  0.10145
## Year2007         -0.0775      0.2035     -0.38  0.70726
## Year2008          0.0740      0.2470      0.30  0.76727
## Year2010         -0.8589      0.1816     -4.73  0.00011 ***
## Year2012          0.6273      0.3145      1.99  0.05920 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.691, Adjusted R-squared:  0.396
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 35 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.780  0.943  0.968  0.956  0.996  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      2.38e-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 2.498e+00 1      1.58
## LastAuthorFemale -3.139e+14 1      NaN
## Year             -7.847e+14 14      NaN

```

Residuals from first and last author



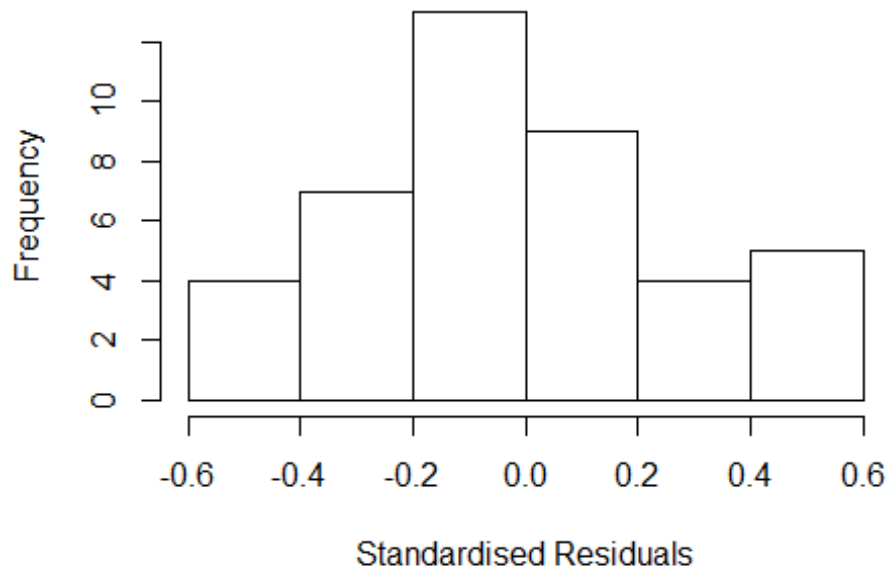
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -5.24e-01 -1.76e-01 -7.22e-16 1.63e-01 5.81e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09224 0.18379 5.94 3.3e-06 ***
## FirstAuthorFemale1 0.00786 0.14271 0.06 0.95652
## LastAuthorFemale1 0.20076 0.18379 1.09 0.28509
## Year1997 0.40988 0.65215 0.63 0.53537
## Year1998 0.35195 0.51038 0.69 0.49681
## Year1999 0.32376 0.18379 1.76 0.09036 .
## Year2000 0.06183 0.20158 0.31 0.76160
## Year2001 1.09476 0.18379 5.96 3.2e-06 ***
## Year2002 0.23876 0.18379 1.30 0.20576
## Year2003 0.18326 0.25308 0.72 0.47570
## Year2004 0.36761 0.21981 1.67 0.10692
## Year2005 1.04226 0.19723 5.28 1.8e-05 ***
```

```

## Year2006          0.10462    0.15509    0.67  0.50614
## Year2007          -0.12107    0.20986   -0.58  0.56916
## Year2008           0.08688    0.16254    0.53  0.59770
## Year2010          -0.80524    0.18379   -4.38  0.00019 ***
## Year2012           0.62783    0.21693    2.89  0.00777 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.327
## Multiple R-squared:  0.618, Adjusted R-squared:  0.374
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 32 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.734  0.878  0.960  0.922  0.981  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      2.38e-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 3.779 1          1.944
## Year              3.779 14          1.049

```


Residuals from first author



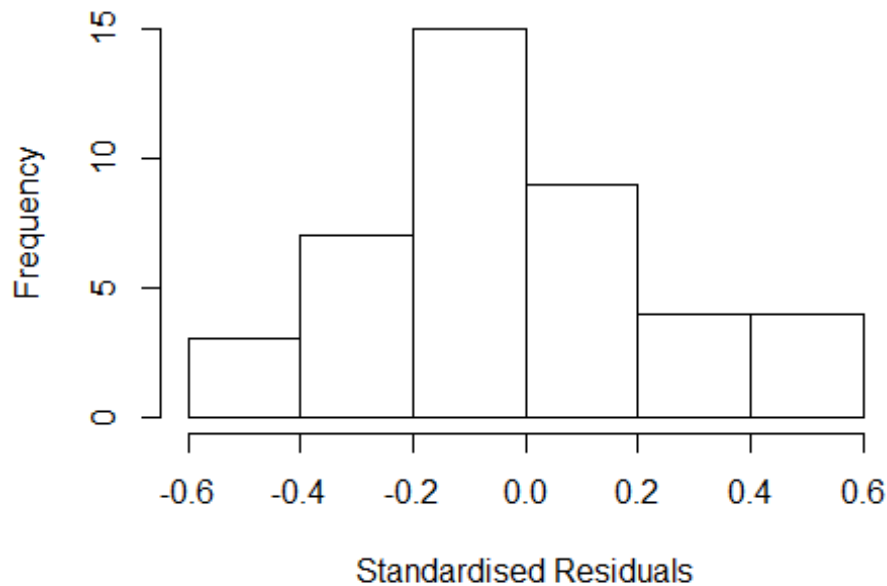
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.92e-01 -1.88e-01  6.66e-16  1.59e-01  5.93e-01
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.29e+00   2.80e-08  4.61e+07 < 2e-16 ***
## FirstAuthorFemale1 2.10e-02   1.56e-01  1.30e-01  0.8944
## Year1997         3.09e-01   7.59e-01  4.10e-01  0.6867
## Year1998         2.45e-01   2.88e-01  8.50e-01  0.4019
## Year1999         1.23e-01   2.91e-08  4.23e+06 < 2e-16 ***
## Year2000        -1.08e-01   1.31e-01 -8.20e-01  0.4170
## Year2001         8.94e-01   3.06e-08  2.92e+07 < 2e-16 ***
## Year2002         3.80e-02   2.88e-08  1.32e+06 < 2e-16 ***
## Year2003        -1.75e-02   1.69e-01 -1.00e-01  0.9181
## Year2004         2.41e-01   1.55e-01  1.55e+00  0.1326
## Year2005         8.42e-01   7.12e-02  1.18e+01  5.8e-12 ***
## Year2006         1.59e-02   1.45e-01  1.10e-01  0.9130
```

```

## Year2007          -3.26e-01   9.20e-02 -3.55e+00   0.0015 **
## Year2008          -1.35e-02   2.18e-01 -6.00e-02   0.9510
## Year2010          -1.01e+00   2.73e-08 -3.69e+07   < 2e-16 ***
## Year2012          4.21e-01   1.01e-01  4.16e+00   0.0003 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.582, Adjusted R-squared:  0.341
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 34 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.809  0.907  0.971  0.947  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      2.38e-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 9.52e+14 1      3.085e+07
## Year              9.52e+14 14      3.427e+00

```

Residuals from last author



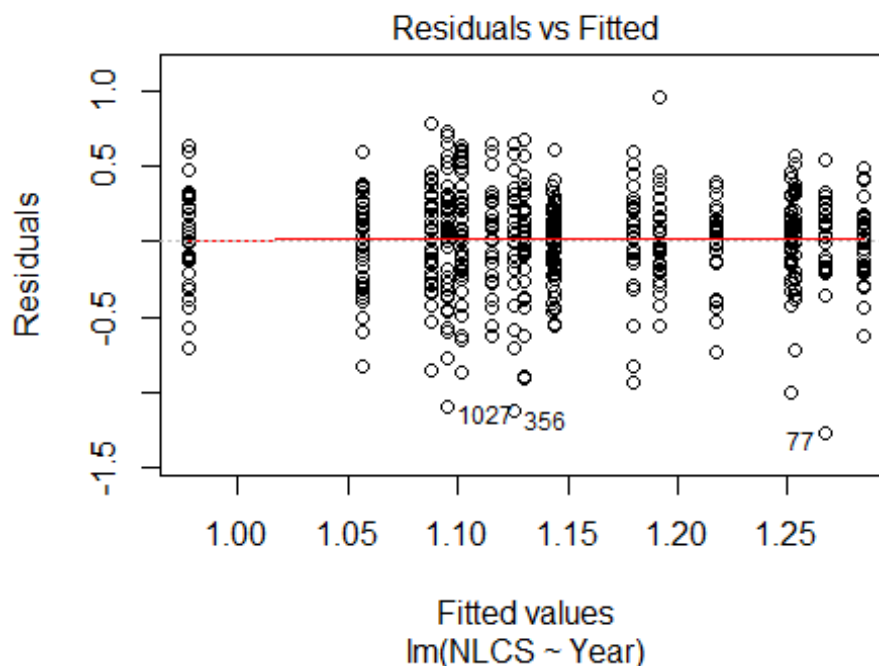
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -5.28e-01 -1.74e-01 -6.94e-17 1.59e-01 5.88e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0872 0.1927 5.64 6.2e-06 ***
## LastAuthorFemale1 0.2058 0.1927 1.07 0.29521
## Year1997 0.4124 0.7237 0.57 0.57366
## Year1998 0.3584 0.5640 0.64 0.53064
## Year1999 0.3288 0.1927 1.71 0.09981 .
## Year2000 0.0679 0.2092 0.32 0.74815
## Year2001 1.0998 0.1927 5.71 5.2e-06 ***
## Year2002 0.2438 0.1927 1.27 0.21692
## Year2003 0.1883 0.2660 0.71 0.48516
## Year2004 0.3787 0.1903 1.99 0.05722 .
## Year2005 1.0473 0.2056 5.09 2.6e-05 ***
## Year2006 0.1026 0.1549 0.66 0.51343
```

```

## Year2007          -0.1136      0.2139    -0.53   0.59984
## Year2008           0.0894      0.1641     0.54   0.59049
## Year2010          -0.8002      0.1927    -4.15   0.00031 ***
## Year2012           0.6368      0.2057     3.10   0.00466 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.309
## Multiple R-squared:  0.624, Adjusted R-squared:  0.406
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 33 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.698  0.865   0.957   0.915   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      2.38e-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: 42"
## [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
##   61   53   48   64   58   47   40   49   45   57   66   57   61   66   51
## 2011 2012
##   60   59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   26   27   35   35   24   26   29   27   38   34   37   33   44   28
## 2011 2012
##   42   33
##

```

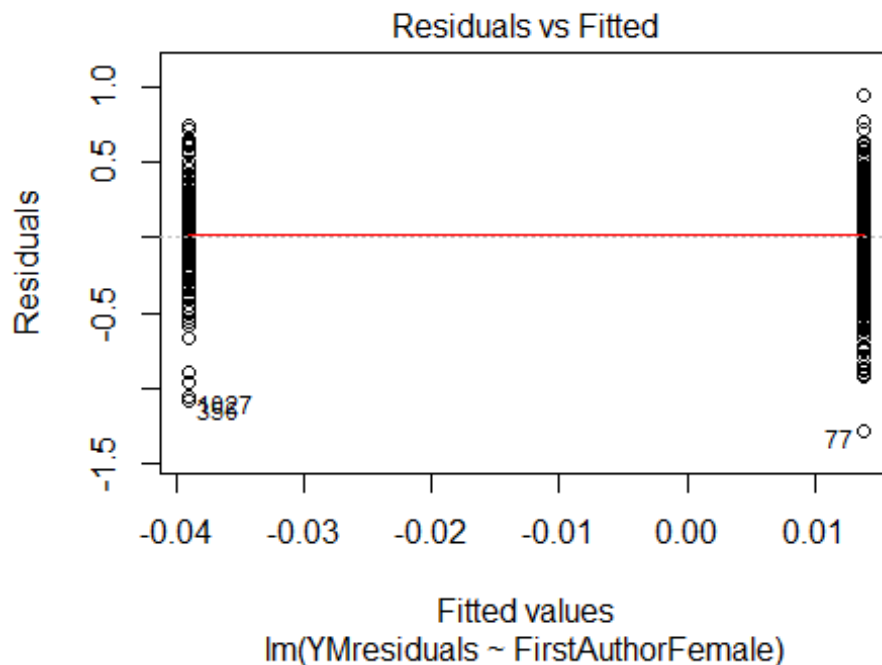
```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    26    26    27    34    33    20    24    27    21    34    31    30    26    38    26
## 2011 2012
##    38    25
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```



```
##
## 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: 4.47769492694043"
## [1] "Male first author team size 2018 geometric mean: 4.14006275927162"
##
## 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 = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.10723250595386"
## [1] "Male last author team size 2018 geometric mean: 4.50072783519754"

## 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 1.194 1          1.093
## LastAuthorFemale  1.149 1          1.072
## UniqueAuthors    2.350 4          1.113
## Year              2.936 16          1.034

## [1] "List of 0 outliers with 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
## -1.34286 -0.18678  0.00491  0.17793  0.95080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.31546    0.07628   17.25 < 2e-16 ***
## FirstAuthorFemale1 -0.04882    0.03708   -1.32  0.1886
## LastAuthorFemale1  0.04508    0.04680    0.96  0.3360
## UniqueAuthors2    -0.04690    0.07088   -0.66  0.5086
## UniqueAuthors3    -0.02416    0.06864   -0.35  0.7250
## UniqueAuthors4    -0.02483    0.07272   -0.34  0.7329
## UniqueAuthors5     0.04272    0.07033    0.61  0.5439
## Year1997           0.02741    0.06607    0.41  0.6785
## Year1998          -0.04367    0.07993   -0.55  0.5851
## Year1999          -0.11644    0.06583   -1.77  0.0776 .
## Year2000          -0.00513    0.06052   -0.08  0.9324
## Year2001          -0.11697    0.11614   -1.01  0.3144
## Year2002          -0.04355    0.08323   -0.52  0.6010
## Year2003          -0.04227    0.06526   -0.65  0.5175
## Year2004          -0.11579    0.06187   -1.87  0.0619 .
## Year2005          -0.11572    0.06465   -1.79  0.0741 .
## Year2006          -0.10333    0.08196   -1.26  0.2081
## Year2007          -0.26731    0.07346   -3.64  0.0003 ***
## Year2008          -0.21062    0.07447   -2.83  0.0049 **
## Year2009          -0.16031    0.07868   -2.04  0.0422 *
## Year2010          -0.20312    0.08453   -2.40  0.0167 *
## Year2011          -0.18543    0.07861   -2.36  0.0187 *
## Year2012          -0.34187    0.08450   -4.05  6.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.289
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0638
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 30 is an outlier with |weight| <= 0.0002 ( < 0.00021);
## 44 weights are ~= 1. The remaining 441 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.094  0.873   0.952   0.896   0.986   0.999

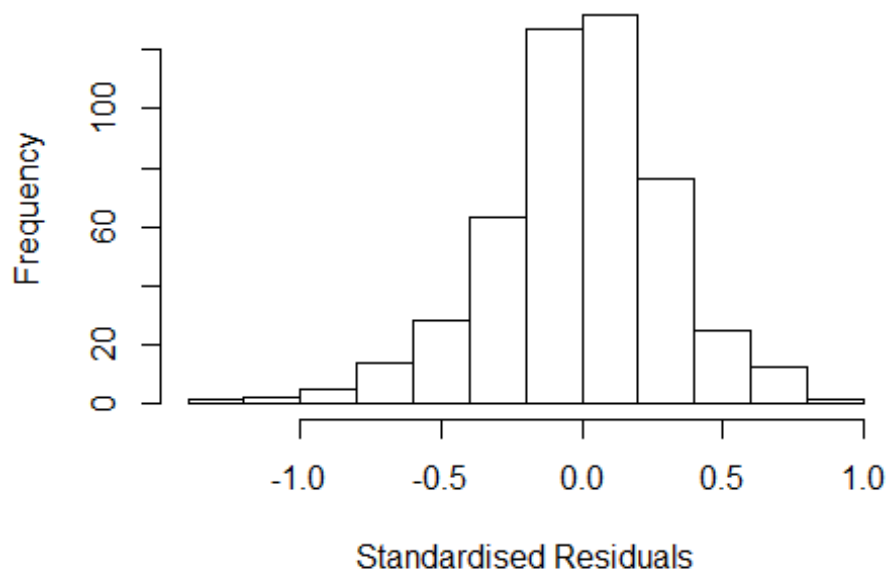
```

```
## 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          2.06e-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"

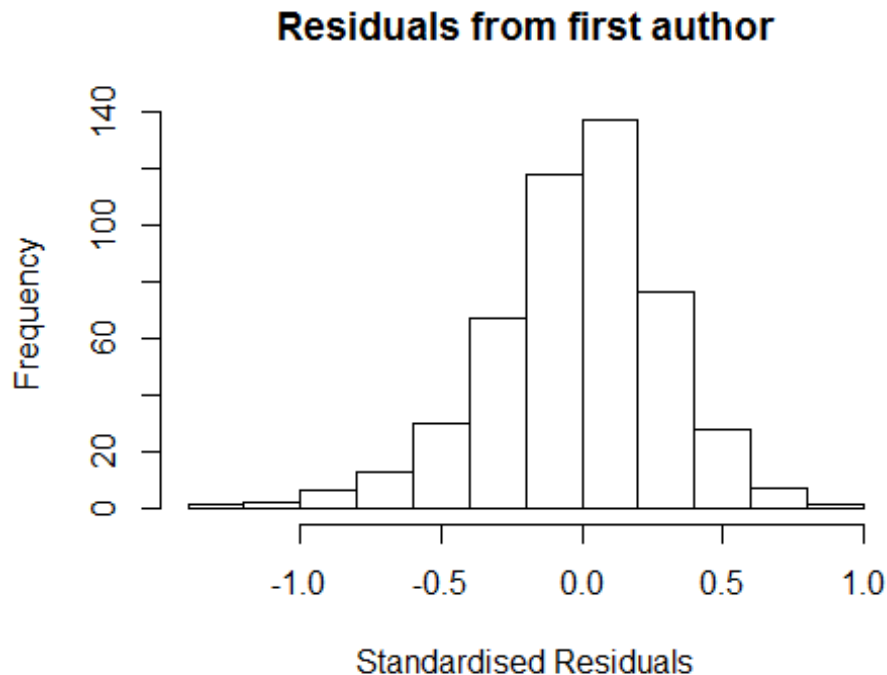
## Warning in lmrob.S(x, y, control = control, mf = mf): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps

## 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
```

Residuals from first and last author and team size




```
## [1] "Regression 3: First author gender, Year as factors"
##               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.139  1      1.067
## Year              1.139 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
## -1.32080 -0.19951  0.00715  0.17201  0.97021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.3023    0.0461   28.23  < 2e-16 ***
## FirstAuthorFemale1 -0.0493    0.0365   -1.35  0.17756
## Year1997         0.0185    0.0621    0.30  0.76528
## Year1998        -0.0402    0.0798   -0.50  0.61504
## Year1999        -0.1245    0.0651   -1.91  0.05648 .
## Year2000        -0.0144    0.0590   -0.24  0.80775
## Year2001        -0.1274    0.1139   -1.12  0.26365
```

```

## Year2002          -0.0410      0.0838   -0.49  0.62459
## Year2003          -0.0386      0.0644   -0.60  0.54941
## Year2004          -0.1198      0.0609   -1.97  0.04982 *
## Year2005          -0.1127      0.0622   -1.81  0.07079 .
## Year2006          -0.1012      0.0796   -1.27  0.20435
## Year2007          -0.2568      0.0747   -3.44  0.00064 ***
## Year2008          -0.2093      0.0750   -2.79  0.00549 **
## Year2009          -0.1481      0.0782   -1.89  0.05889 .
## Year2010          -0.1772      0.0847   -2.09  0.03699 *
## Year2011          -0.1741      0.0774   -2.25  0.02500 *
## Year2012          -0.3296      0.0837   -3.94  9.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.29
## Multiple R-squared:  0.094, Adjusted R-squared:  0.0611
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.003  0.871  0.952  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      2.06e-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 1.13 1          1.063
## Year          1.13 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
## -1.3121 -0.1901  0.0132  0.1769  0.9281
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2940     0.0468   27.64 < 2e-16 ***
## LastAuthorFemale1  0.0479     0.0462    1.04  0.30036
## Year1997          0.0182     0.0642    0.28  0.77742
## Year1998         -0.0430     0.0808   -0.53  0.59466
## Year1999         -0.1220     0.0651   -1.87  0.06153 .
## Year2000         -0.0228     0.0605   -0.38  0.70641
## Year2001         -0.1366     0.1145   -1.19  0.23364
## Year2002         -0.0529     0.0832   -0.63  0.52575
## Year2003         -0.0452     0.0674   -0.67  0.50357
## Year2004         -0.1233     0.0623   -1.98  0.04855 *
## Year2005         -0.1164     0.0636   -1.83  0.06784 .
## Year2006         -0.1142     0.0813   -1.41  0.16064
## Year2007         -0.2625     0.0764   -3.43  0.00065 ***
## Year2008         -0.2258     0.0739   -3.06  0.00236 **
## Year2009         -0.1553     0.0790   -1.97  0.04987 *
## Year2010         -0.1969     0.0848   -2.32  0.02072 *
## Year2011         -0.1981     0.0799   -2.48  0.01353 *
## Year2012         -0.3454     0.0826   -4.18  3.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.292
## Multiple R-squared:  0.0912, Adjusted R-squared:  0.0582
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 445 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.0063  0.8740  0.9560  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          2.06e-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"

```

```

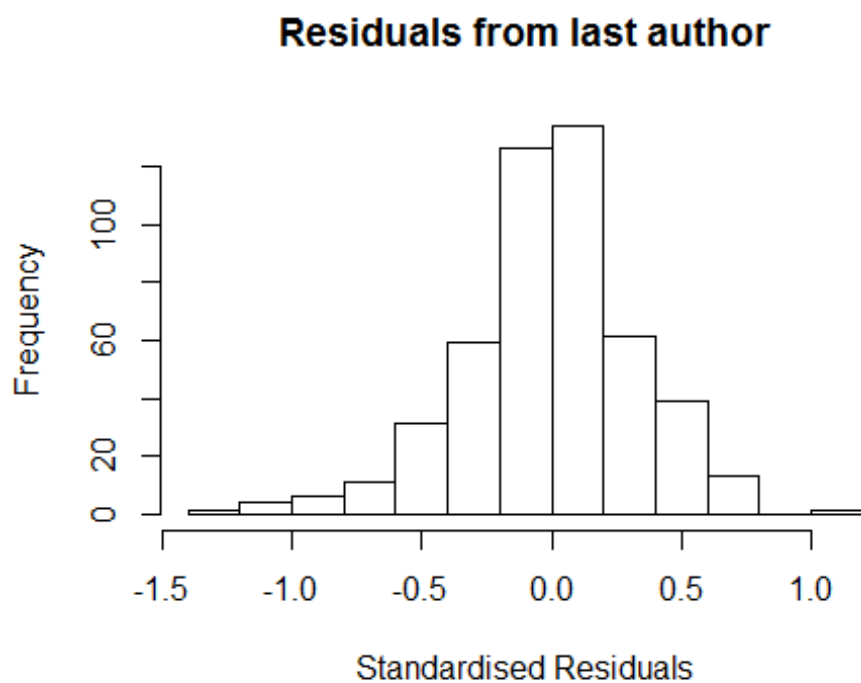
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 486"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1506"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10   9  15   9  21  16  22  17  11  18  37  19  32  27  37
## 2011 2012
##   41  38
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1   2   4   2   6   6   9   9   1   4  20   8   9  18  13
## 2011 2012
##   27  23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1   2   3   2   5   4   6   8   0   4  20   7   6  15  12
## 2011 2012
##   24  22
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.73813722053759"
## [1] "Male first author team size 2018 geometric mean: 3.72625909711739"

## 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 = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.63462649452252"
## [1] "Male last author team size 2018 geometric mean: 3.47672267262469"

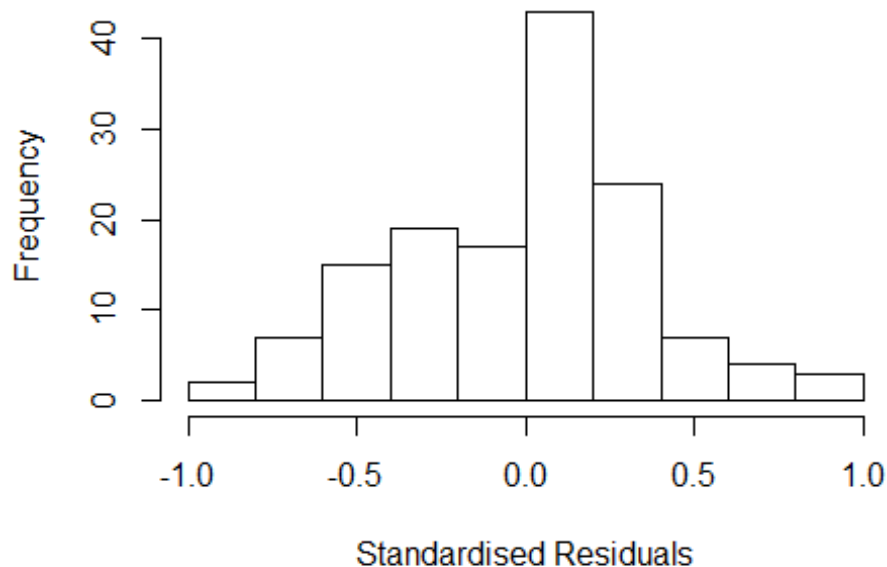
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32, 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 28.766 1      5.363
## LastAuthorFemale  2.181 1      1.477
## UniqueAuthors    2028.264 4      2.591
## Year              6895.691 15     1.343
```

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
## -0.9107 -0.2742 0.0581 0.2254 0.8987
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11e+00 2.48e-08 4.46e+07 < 2e-16 ***
## FirstAuthorFemale1 8.67e-02 9.72e-02 8.90e-01 0.37427
## LastAuthorFemale1 1.59e-01 1.01e-01 1.58e+00 0.11664
## UniqueAuthors2 4.87e-01 1.61e-01 3.03e+00 0.00303 **
## UniqueAuthors3 5.73e-01 1.79e-01 3.20e+00 0.00175 **
## UniqueAuthors4 4.55e-01 1.90e-01 2.39e+00 0.01853 *
## UniqueAuthors5 7.66e-01 1.71e-01 4.47e+00 1.8e-05 ***
## Year1997 -1.56e-01 2.30e-01 -6.80e-01 0.49889
## Year1998 -4.40e-01 2.01e-01 -2.19e+00 0.03079 *
## Year1999 1.60e-01 1.73e-01 9.20e-01 0.35913
```

```

## Year2000      -8.77e-01    2.75e-01 -3.19e+00    0.00184 **
## Year2001      -7.06e-01    2.58e-01 -2.74e+00    0.00711 **
## Year2002      -3.80e-01    2.88e-01 -1.32e+00    0.18927 .
## Year2003      -4.05e-01    2.15e-01 -1.88e+00    0.06276 .
## Year2005      -8.21e-01    1.67e-01 -4.93e+00    2.7e-06 ***
## Year2006      -6.00e-01    1.46e-01 -4.11e+00    7.2e-05 ***
## Year2007      -3.75e-01    1.91e-01 -1.96e+00    0.05236 .
## Year2008      -4.69e-01    1.87e-01 -2.50e+00    0.01377 *
## Year2009      -3.43e-01    1.65e-01 -2.08e+00    0.03981 *
## Year2010      -2.60e-01    1.85e-01 -1.40e+00    0.16334
## Year2011      -5.16e-01    1.77e-01 -2.92e+00    0.00413 **
## Year2012      -6.64e-01    1.70e-01 -3.91e+00    0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.39,   Adjusted R-squared:  0.283
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 130 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.563  0.887  0.959  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           7.09e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 24.947 1 4.995
## LastAuthorFemale 1.632 1 1.277
## Year 40.350 15 1.131
## [1] "List of 0 outliers with 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
## -1.2295 -0.2811  0.0316  0.2512  0.9231
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.11e+00   5.57e-08  1.99e+07 < 2e-16 ***
## FirstAuthorFemale1  1.91e-01   9.09e-02  2.10e+00  0.0377 *
## LastAuthorFemale1  1.88e-01   9.37e-02  2.01e+00  0.0468 *
## Year1997         3.55e-02   4.36e-01  8.00e-02  0.9352
## Year1998         9.56e-02   1.33e-01  7.20e-01  0.4726
## Year1999         5.10e-01   9.29e-02  5.49e+00  2.2e-07 ***
## Year2000        -3.54e-01   1.80e-01 -1.97e+00  0.0514 .
## Year2001        -2.39e-01   2.51e-01 -9.50e-01  0.3439
## Year2002         1.22e-01   2.59e-01  4.70e-01  0.6400
## Year2003        -1.22e-01   1.79e-01 -6.80e-01  0.4961
## Year2005        -6.27e-01   2.68e-01 -2.34e+00  0.0209 *
## Year2006        -2.56e-01   8.80e-02 -2.91e+00  0.0042 **
## Year2007         1.66e-01   1.59e-01  1.04e+00  0.2987
## Year2008         8.52e-02   1.28e-01  6.60e-01  0.5077
## Year2009         1.98e-01   1.04e-01  1.90e+00  0.0600 .
## Year2010         2.57e-01   1.01e-01  2.55e+00  0.0121 *
## Year2011         8.56e-02   1.06e-01  8.10e-01  0.4216
## Year2012        -1.00e-01   1.31e-01 -7.70e-01  0.4443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.245, Adjusted R-squared:  0.141
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 128 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.364  0.882  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      7.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd

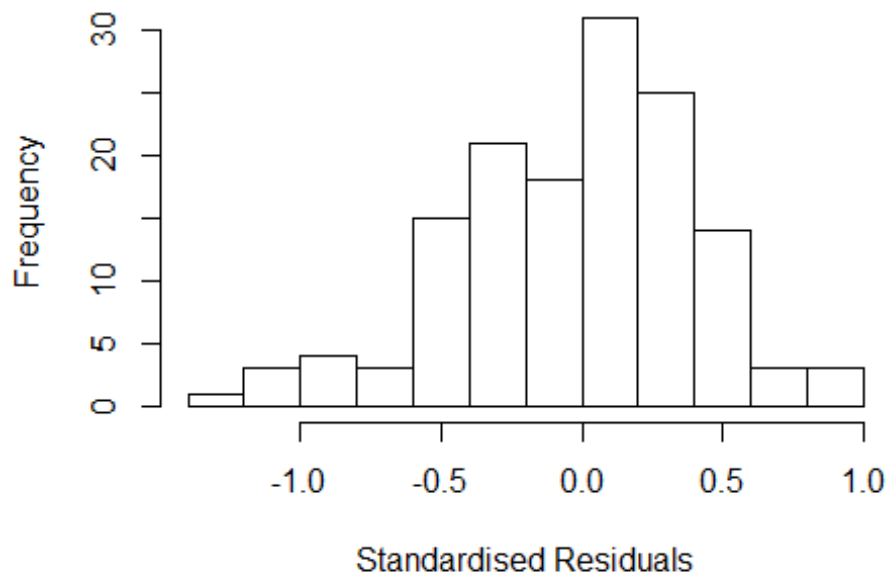
```



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

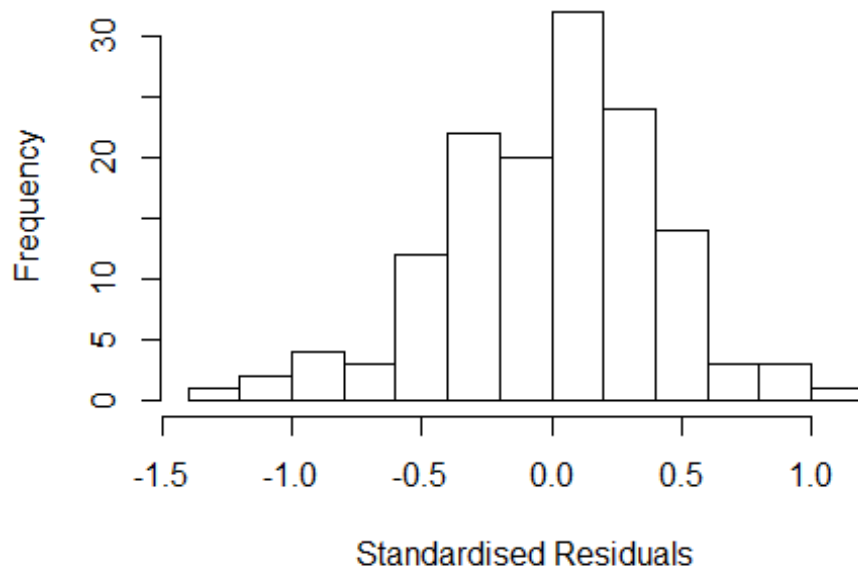
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```

Residuals from first and last author



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 26.57 1           5.155
## Year              26.57 15          1.116
```

Residuals from first author



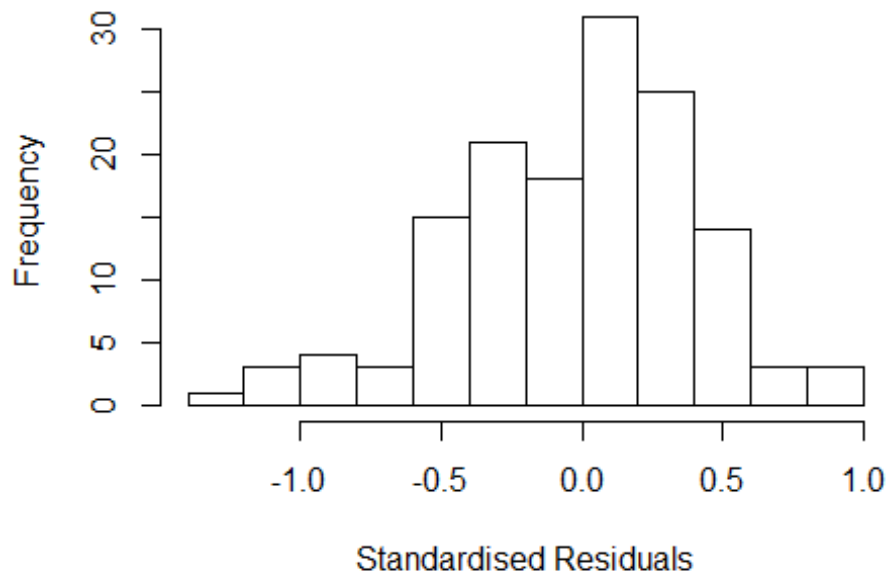
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2259 -0.3144 0.0265 0.2755 1.0525
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1080 0.0000 Inf < 2e-16 ***
## FirstAuthorFemale1 0.2170 0.0941 2.31 0.0228 *
## Year1997 0.0225 0.4200 0.05 0.9574
## Year1998 0.0957 0.1328 0.72 0.4727
## Year1999 0.4840 0.0961 5.04 1.6e-06 ***
## Year2000 -0.3436 0.2058 -1.67 0.0975 .
## Year2001 -0.2551 0.2534 -1.01 0.3159
## Year2002 0.1179 0.2594 0.45 0.6501
## Year2003 -0.0754 0.1897 -0.40 0.6919
## Year2005 -0.6388 0.2630 -2.43 0.0166 *
## Year2006 -0.2037 0.0812 -2.51 0.0135 *
## Year2007 0.2272 0.1760 1.29 0.1991
```

```

## Year2008          0.0724      0.1340      0.54      0.5902
## Year2009          0.2265      0.0956      2.37      0.0194 *
## Year2010          0.3062      0.0997      3.07      0.0026 **
## Year2011          0.1326      0.1030      1.29      0.2004
## Year2012         -0.0415      0.1349     -0.31      0.7586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.223, Adjusted R-squared:  0.123
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 127 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.355  0.879  0.946  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      7.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.667 1          1.291
## Year            1.667 15          1.017

```

Residuals from last author



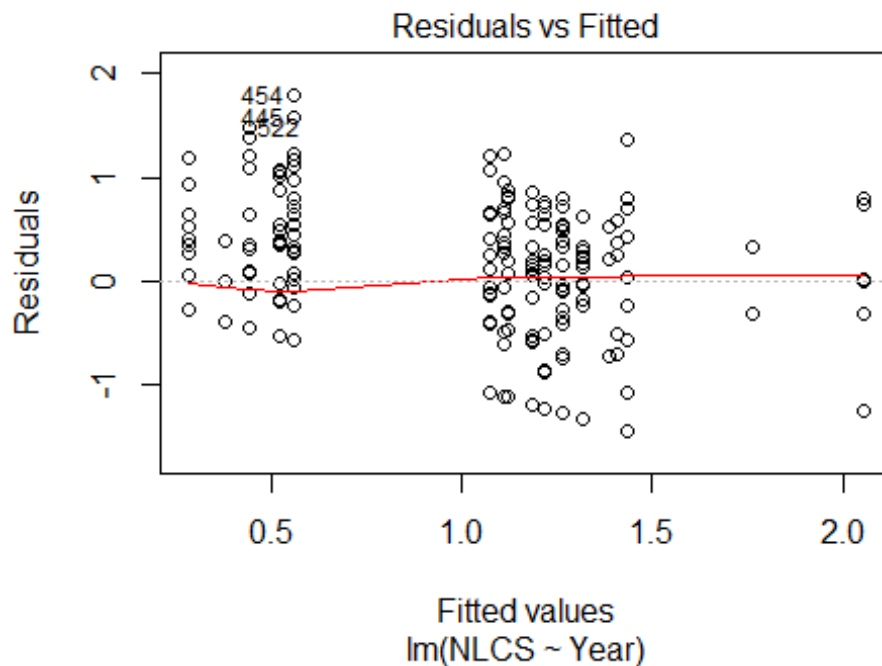
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2864 -0.3006 0.0469 0.2648 0.9935
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11e+00 5.16e-08 2.15e+07 <2e-16 ***
## LastAuthorFemale1 2.16e-01 9.79e-02 2.21e+00 0.0290 *
## Year1997 1.31e-01 6.25e-01 2.10e-01 0.8344
## Year1998 9.56e-02 1.33e-01 7.20e-01 0.4726
## Year1999 7.01e-01 1.91e-02 3.67e+01 <2e-16 ***
## Year2000 -3.43e-01 2.06e-01 -1.67e+00 0.0979 .
## Year2001 -1.24e-01 2.33e-01 -5.30e-01 0.5940
## Year2002 1.78e-01 2.62e-01 6.80e-01 0.4974
## Year2003 -5.58e-02 2.10e-01 -2.70e-01 0.7909
## Year2005 -5.60e-01 3.15e-01 -1.78e+00 0.0779 .
## Year2006 -1.70e-01 8.07e-02 -2.11e+00 0.0369 *
## Year2007 1.57e-01 1.58e-01 9.90e-01 0.3220
```

```

## Year2008          1.80e-01   8.90e-02   2.02e+00   0.0454 *
## Year2009          1.95e-01   1.06e-01   1.84e+00   0.0684 .
## Year2010          2.84e-01   1.01e-01   2.80e+00   0.0059 **
## Year2011          1.25e-01   1.05e-01   1.19e+00   0.2355
## Year2012         -2.59e-02   1.14e-01  -2.30e-01   0.8212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.22,   Adjusted R-squared:  0.12
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 130 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.313  0.867  0.953  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      7.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 141"
## [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
##   23   33   53   38   41   62   61   82   49   41   56   51   66   52   45
## 2011 2012
##   32   37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    3    3    2    5    9   39   47   31   13   17   26   11   16   15
## 2011 2012
##   11   14

```

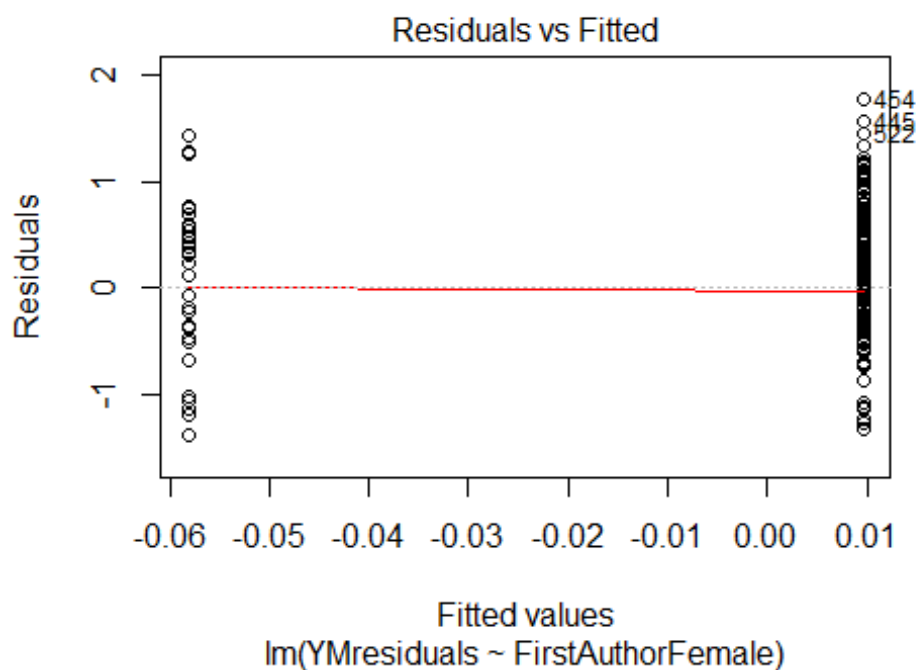
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    3    3    2    3    7   34   44   30    9   14   24    8   13   13
## 2011 2012
##   11   13
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity 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 = 2.9, df = 1, p-value = 0.09
## [1] "Female first author team size 2018 geometric mean: 3.86825415074099"
## [1] "Male first author team size 2018 geometric mean: 2.57438645139409"
## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3"
## [1] "Male last author team size 2018 geometric mean: 2.86768465972161"

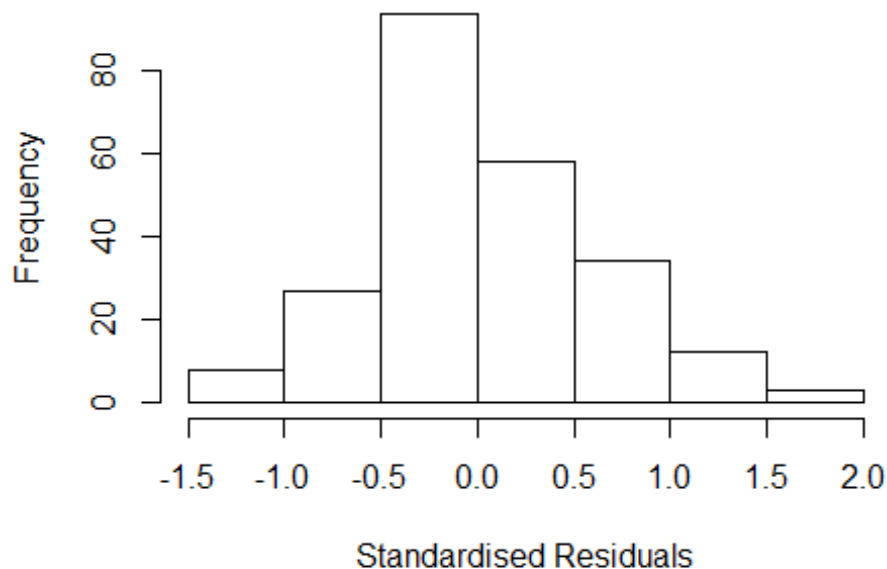
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVI	F	Df	GVI ^{1/(2*Df)}
FirstAuthorFemale	2.382	1		1.543
LastAuthorFemale	1.751	1		1.323
UniqueAuthors	9.366	4		1.323
Year	14.329	16		1.087

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
## -1.4529 -0.3451 -0.0645 0.4019 1.7923
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.07624 0.49118 4.23 3.5e-05 ***
## FirstAuthorFemale1 -0.03910 0.18745 -0.21 0.83495
## LastAuthorFemale1 -0.01098 0.13630 -0.08 0.93584
## UniqueAuthors2 0.21779 0.11896 1.83 0.06853 .
## UniqueAuthors3 0.24981 0.13249 1.89 0.06072 .
## UniqueAuthors4 0.11655 0.21066 0.55 0.58068
## UniqueAuthors5 0.00808 0.20181 0.04 0.96811
## Year1997 -0.68772 0.68277 -1.01 0.31495
## Year1998 -1.84912 0.50572 -3.66 0.00032 ***
## Year1999 -0.49692 0.59512 -0.83 0.40466
```

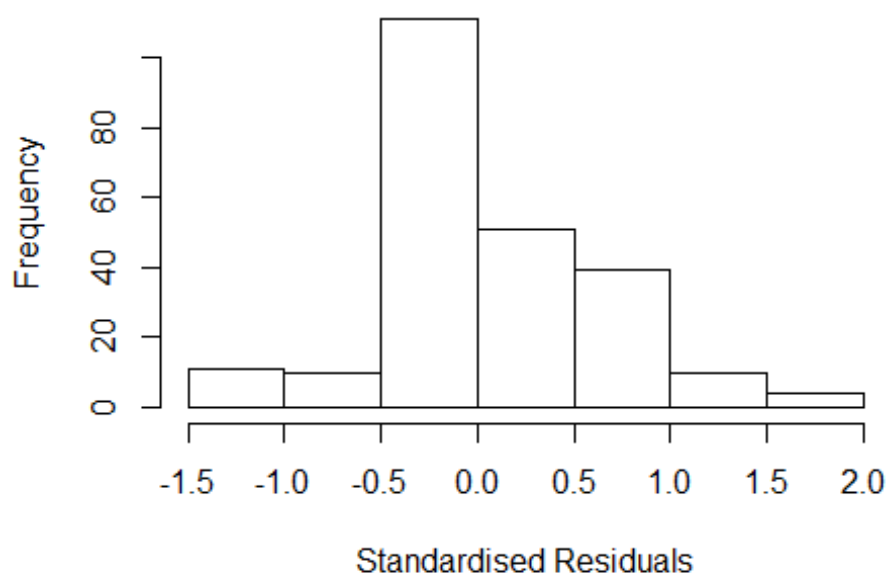


```

## Year2000      -0.98187      0.57402      -1.71      0.08863 .
## Year2001      -0.80083      1.01263      -0.79      0.42991
## Year2002      -1.91332      0.51463      -3.72      0.00026 ***
## Year2003      -1.73932      0.52128      -3.34      0.00100 **
## Year2004      -1.85530      0.51986      -3.57      0.00044 ***
## Year2005      -1.16790      0.56698      -2.06      0.04063 *
## Year2006      -0.98174      0.53059      -1.85      0.06566 .
## Year2007      -1.73118      0.51881      -3.34      0.00100 ***
## Year2008      -1.16658      0.54568      -2.14      0.03367 *
## Year2009      -1.26183      0.57092      -2.21      0.02816 *
## Year2010      -1.06068      0.56396      -1.88      0.06137 .
## Year2011      -0.89771      0.51567      -1.74      0.08315 .
## Year2012      -0.91066      0.54673      -1.67      0.09725 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.443, Adjusted R-squared:  0.386
## Convergence in 34 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 219 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.181  0.855  0.948  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          4.24e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.897 1          1.377
## LastAuthorFemale  1.486 1          1.219
## Year              2.067 16          1.023

```

Residuals from first and last author



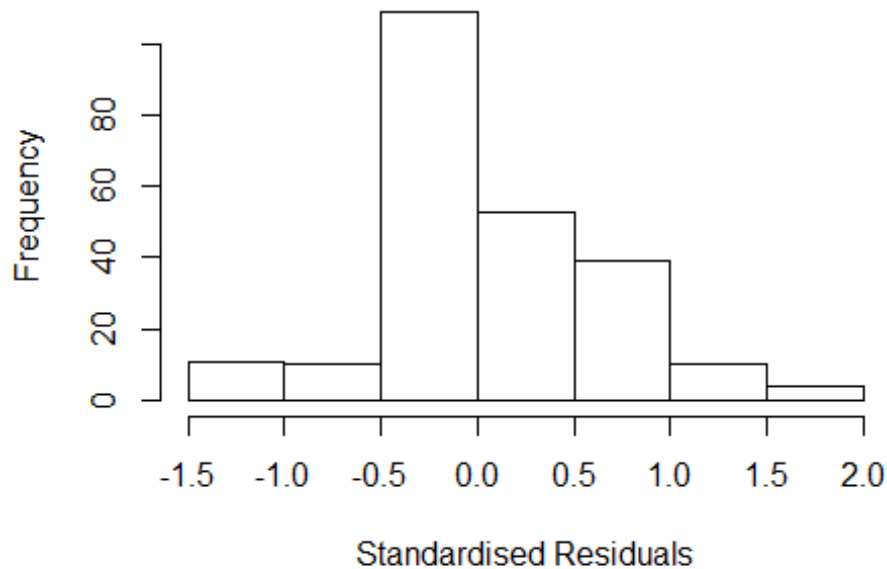
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4025 -0.3469 -0.0692 0.4313 1.9266
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.17095 0.43733 4.96 1.4e-06 ***
## FirstAuthorFemale1 -0.07817 0.17818 -0.44 0.66131
## LastAuthorFemale1 0.00146 0.12733 0.01 0.99084
## Year1997 -0.74655 0.56338 -1.33 0.18652
## Year1998 -1.78981 0.47429 -3.77 0.00021 ***
## Year1999 -0.40845 0.50191 -0.81 0.41666
## Year2000 -1.02800 0.59401 -1.73 0.08494 .
## Year2001 -0.69031 1.01250 -0.68 0.49610
## Year2002 -1.89414 0.45031 -4.21 3.8e-05 ***
## Year2003 -1.75057 0.46426 -3.77 0.00021 ***
## Year2004 -1.82408 0.46416 -3.93 0.00011 ***
## Year2005 -1.12101 0.50288 -2.23 0.02683 *
```

```

## Year2006      -0.88854    0.47386   -1.88  0.06212 .
## Year2007      -1.70415    0.46785   -3.64  0.00034 ***
## Year2008      -1.10408    0.49087   -2.25  0.02550 *
## Year2009      -1.22021    0.55199   -2.21  0.02811 *
## Year2010      -0.96169    0.49855   -1.93  0.05504 .
## Year2011      -0.78393    0.44688   -1.75  0.08080 .
## Year2012      -0.78890    0.47671   -1.65  0.09939 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.5
## Multiple R-squared:  0.436, Adjusted R-squared:  0.389
## Convergence in 35 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 219 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  0.860  0.937  0.876  0.974  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.482 1      1.218
## Year              1.482 16      1.012

```

Residuals from first author



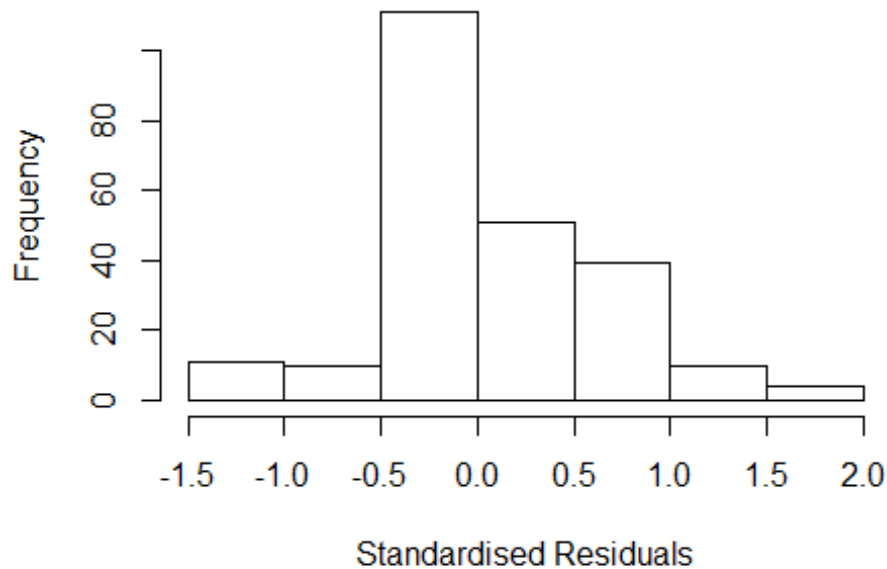
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3930 -0.3515 -0.0677 0.4276 1.9224
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.1659 0.4296 5.04 9.7e-07 ***
## FirstAuthorFemale1 -0.0808 0.1572 -0.51 0.60787
## Year1997 -0.7430 0.5601 -1.33 0.18604
## Year1998 -1.7842 0.4697 -3.80 0.00019 ***
## Year1999 -0.4034 0.4949 -0.82 0.41589
## Year2000 -1.0204 0.5860 -1.74 0.08302 .
## Year2001 -0.6921 0.9583 -0.72 0.47098
## Year2002 -1.8880 0.4434 -4.26 3.1e-05 ***
## Year2003 -1.7413 0.4568 -3.81 0.00018 ***
## Year2004 -1.8144 0.4559 -3.98 9.4e-05 ***
## Year2005 -1.1146 0.4966 -2.24 0.02582 *
## Year2006 -0.8845 0.4652 -1.90 0.05861 .
```

```

## Year2007          -1.6962      0.4626   -3.67  0.00031 ***
## Year2008          -1.0997      0.4836   -2.27  0.02394 *
## Year2009          -1.2141      0.5433   -2.23  0.02645 *
## Year2010          -0.9589      0.4924   -1.95  0.05274 .
## Year2011          -0.7810      0.4401   -1.77  0.07733 .
## Year2012          -0.7860      0.4682   -1.68  0.09461 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.508
## Multiple R-squared:  0.431, Adjusted R-squared:  0.386
## Convergence in 33 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 219 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.120  0.863   0.937   0.880   0.975   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.262 1          1.123
## Year            1.262 16          1.007

```

Residuals from last author



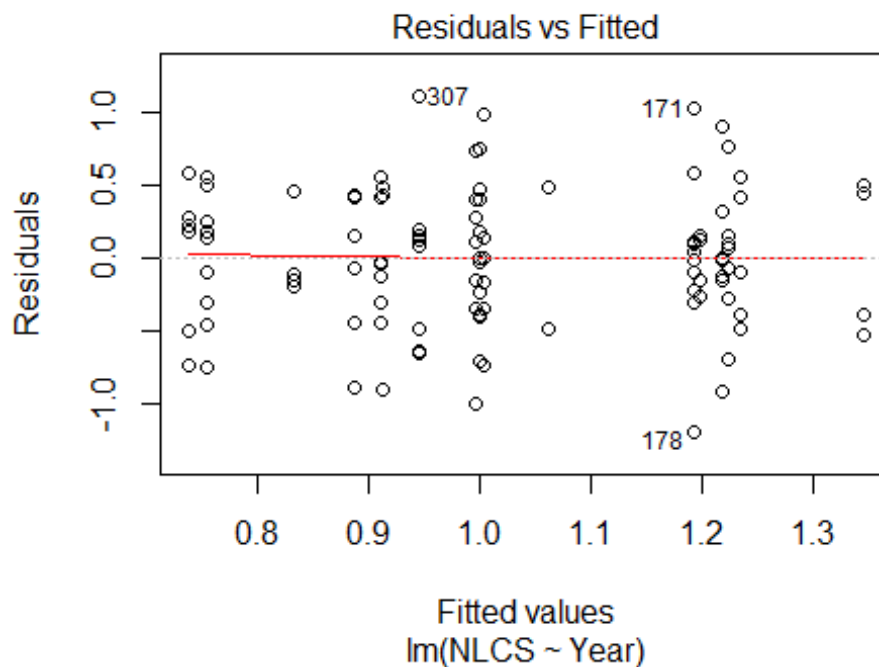
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5154 -0.3357 -0.0709 0.4313 1.9362
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.1876 0.4425 4.94 1.5e-06 ***
## LastAuthorFemale1 -0.0316 0.1168 -0.27 0.78705
## Year1997 -0.7623 0.5673 -1.34 0.18044
## Year1998 -1.7941 0.4877 -3.68 0.00030 ***
## Year1999 -0.4251 0.5066 -0.84 0.40234
## Year2000 -1.0460 0.5996 -1.74 0.08250 .
## Year2001 -0.6721 1.0719 -0.63 0.53127
## Year2002 -1.9173 0.4509 -4.25 3.1e-05 ***
## Year2003 -1.7767 0.4636 -3.83 0.00017 ***
## Year2004 -1.8519 0.4628 -4.00 8.6e-05 ***
## Year2005 -1.1292 0.5143 -2.20 0.02917 *
## Year2006 -0.9140 0.4731 -1.93 0.05465 .
```

```

## Year2007          -1.7261      0.4703    -3.67  0.00030 ***
## Year2008          -1.1203      0.4958    -2.26  0.02484 *
## Year2009          -1.2341      0.5609    -2.20  0.02883 *
## Year2010          -0.9804      0.5053    -1.94  0.05363 .
## Year2011          -0.7994      0.4514    -1.77  0.07795 .
## Year2012          -0.8218      0.4659    -1.76  0.07913 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.495
## Multiple R-squared:  0.441, Adjusted R-squared:  0.398
## Convergence in 34 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 221 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0924 0.8640 0.9380 0.8750 0.9730 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 236"
## [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
##   10   6  17  10  17  17  11  17   9  15  19  25  19  12  22
## 2011 2012
##   17  23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2   4   5   4   7   7   3   9   6   7  10  10   7   5   7
## 2011 2012

```

```
##      9      7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      2      3      3      4      5      5      2      8      5      6     10     10      6      3      6
## 2011 2012
##      6      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 = 8.4, df = 16, p-value = 0.9
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.069, df = 1, p-value = 0.8

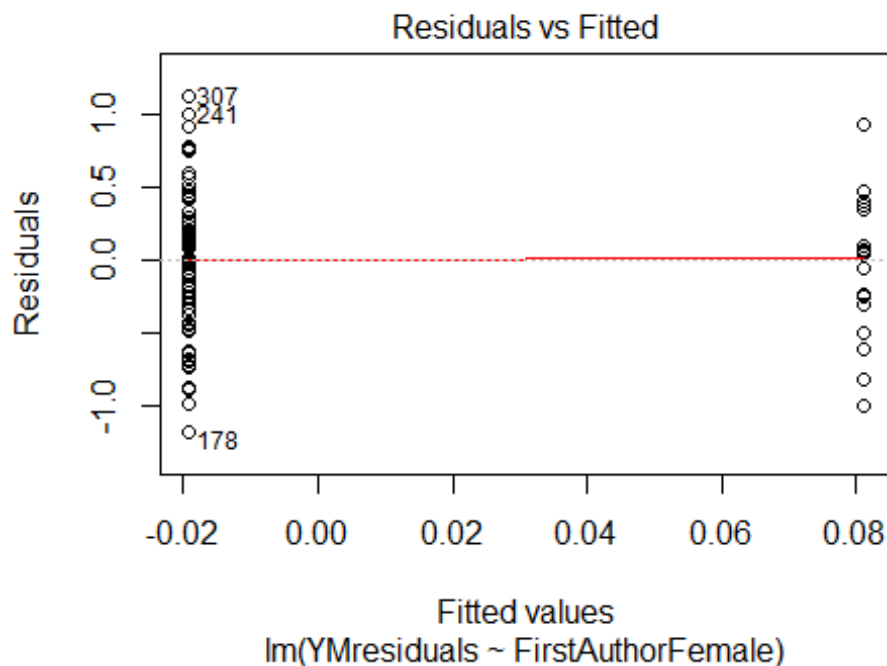
## [1] "Female first author team size 2018 geometric mean: 3.92329869898433"
## [1] "Male first author team size 2018 geometric mean: 3.46393924115679"

## 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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.7062880231628"
## [1] "Male last author team size 2018 geometric mean: 3.5822358264556"

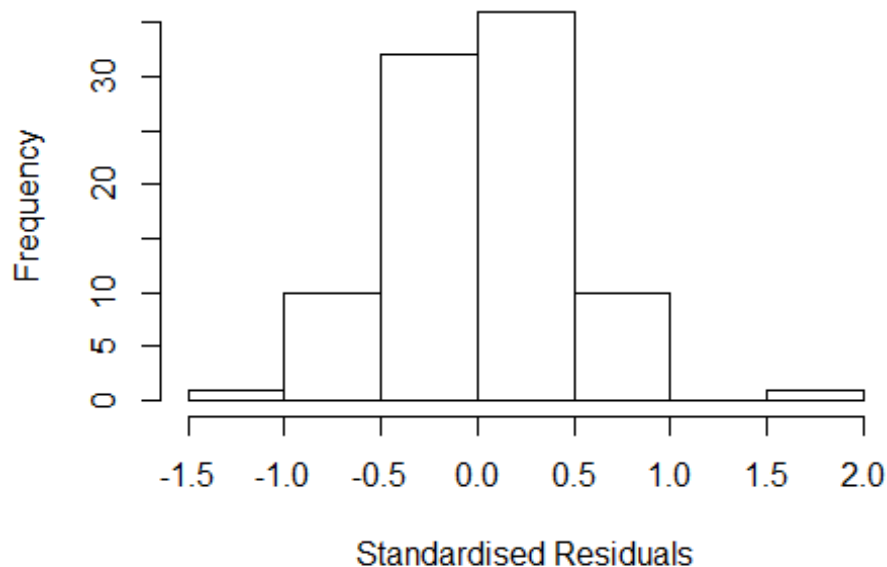
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 76, 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	2.178	1	1.476
LastAuthorFemale	2.340	1	1.530
UniqueAuthors	60.708	4	1.671
Year	109.531	16	1.158

Residuals from first and last author and team size



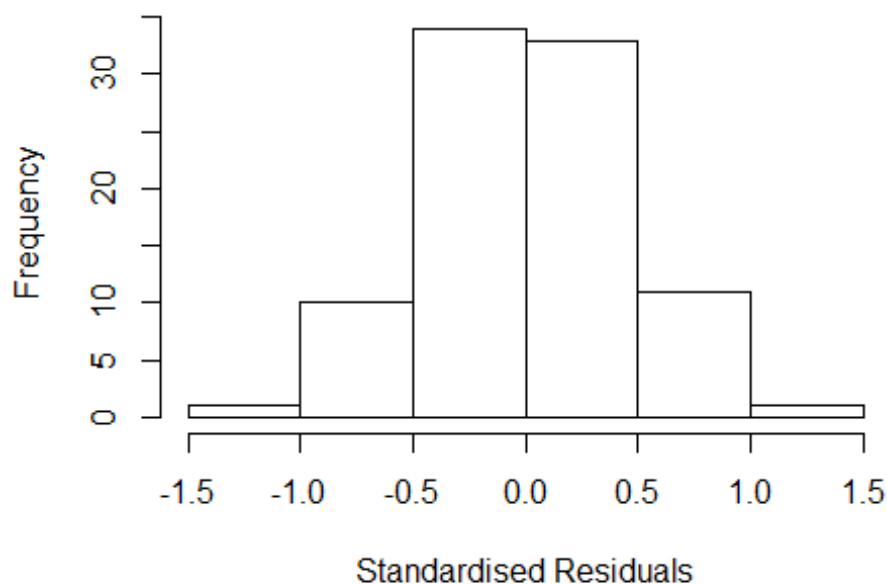
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0238 -0.2784 0.0194 0.2650 1.5378
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6109 0.3547 1.72 0.08963 .
## FirstAuthorFemale1 0.2195 0.1526 1.44 0.15504
## LastAuthorFemale1 -0.0818 0.2161 -0.38 0.70629
## UniqueAuthors2 0.3269 0.2303 1.42 0.16043
## UniqueAuthors3 0.6064 0.1737 3.49 0.00086 ***
## UniqueAuthors4 0.3548 0.2073 1.71 0.09172 .
## UniqueAuthors5 0.5957 0.2525 2.36 0.02125 *
## Year1997 -0.0516 0.3355 -0.15 0.87828
## Year1998 -0.0457 0.3528 -0.13 0.89727
## Year1999 0.1946 0.4508 0.43 0.66742
```

```

## Year2000          0.1270      0.3390      0.37  0.70914
## Year2001          0.1166      0.3765      0.31  0.75776
## Year2002         -0.4065      1.2935     -0.31  0.75428
## Year2003         -0.0971      0.3431     -0.28  0.77810
## Year2004         -0.4186      0.3664     -1.14  0.25733
## Year2005         -0.0937      0.3469     -0.27  0.78794
## Year2006          0.0581      0.3777      0.15  0.87819
## Year2007         -0.1356      0.3420     -0.40  0.69311
## Year2008         -0.1557      0.4343     -0.36  0.72100
## Year2009          0.2588      0.3610      0.72  0.47592
## Year2010          0.2321      0.3683      0.63  0.53078
## Year2011         -0.2631      0.3767     -0.70  0.48737
## Year2012         -0.2219      0.4462     -0.50  0.62065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.294, Adjusted R-squared:  0.0619
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 86 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.886  0.963   0.911   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      1.11e-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 1.447 1      1.203
## LastAuthorFemale  2.019 1      1.421
## Year              2.714 16      1.032

```

Residuals from first and last author



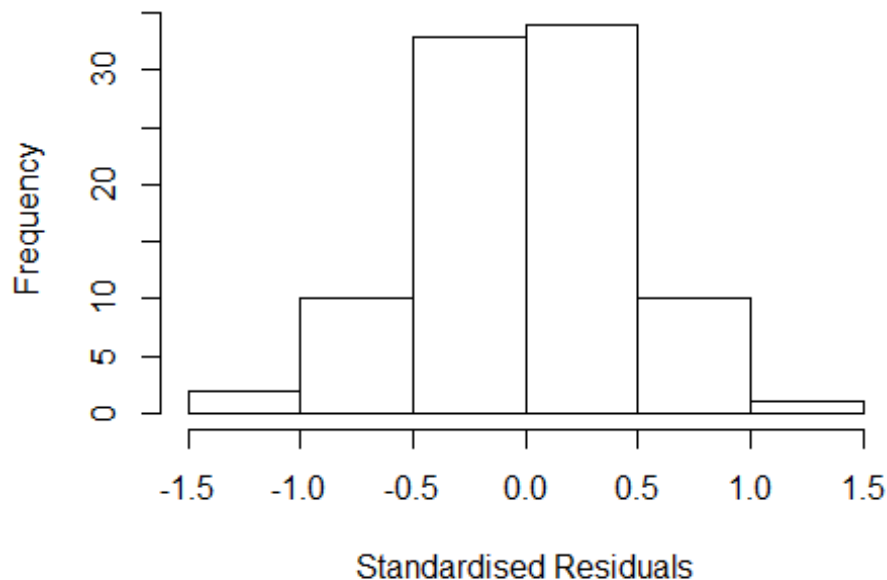
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18419 -0.33940 0.00411 0.37433 1.08305
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9654 0.3116 3.10 0.0028 **
## FirstAuthorFemale1 0.1922 0.1396 1.38 0.1728
## LastAuthorFemale1 -0.1326 0.2085 -0.64 0.5268
## Year1997 -0.1021 0.3581 -0.29 0.7764
## Year1998 0.2085 0.3107 0.67 0.5045
## Year1999 0.2398 0.3856 0.62 0.5359
## Year2000 0.1055 0.3328 0.32 0.7523
## Year2001 -0.0204 0.3634 -0.06 0.9554
## Year2002 -0.2299 0.6539 -0.35 0.7262
## Year2003 -0.2087 0.3564 -0.59 0.5600
## Year2004 -0.2743 0.3963 -0.69 0.4911
## Year2005 -0.1128 0.3724 -0.30 0.7628
```

```

## Year2006          0.2188      0.3561      0.61      0.5410
## Year2007          -0.0168      0.3374     -0.05      0.9604
## Year2008          -0.0555      0.3988     -0.14      0.8898
## Year2009           0.2261      0.3581      0.63      0.5297
## Year2010           0.2093      0.3951      0.53      0.5980
## Year2011          -0.2075      0.3641     -0.57      0.5706
## Year2012           0.0349      0.3988      0.09      0.9306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.158, Adjusted R-squared:  -0.0554
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 80 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.646  0.914  0.958  0.936  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      1.11e-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 1.371 1          1.171
## Year              1.371 16          1.010

```

Residuals from first author



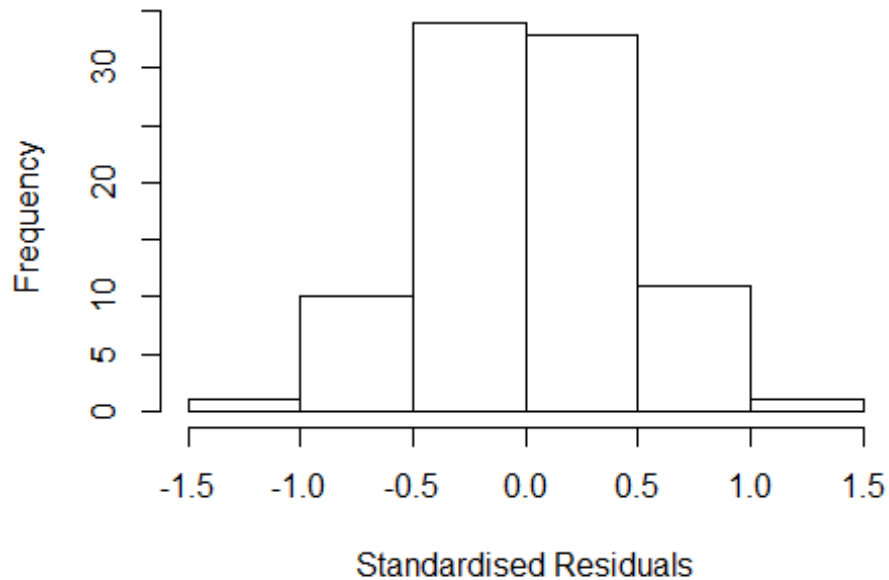
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15513 -0.32961 0.00745 0.37529 1.10485
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97376 0.31782 3.06 0.0031 **
## FirstAuthorFemale1 0.17549 0.13739 1.28 0.2056
## Year1997 -0.11023 0.36342 -0.30 0.7625
## Year1998 0.20573 0.31707 0.65 0.5185
## Year1999 0.24335 0.39122 0.62 0.5359
## Year2000 0.07383 0.34533 0.21 0.8313
## Year2001 -0.02865 0.36848 -0.08 0.9382
## Year2002 -0.36250 0.62987 -0.58 0.5667
## Year2003 -0.23080 0.35837 -0.64 0.5216
## Year2004 -0.28303 0.40043 -0.71 0.4820
## Year2005 -0.11892 0.37777 -0.31 0.7538
## Year2006 0.18137 0.35217 0.52 0.6081
```

```

## Year2007          -0.03184    0.34086   -0.09    0.9258
## Year2008          -0.08561    0.39593   -0.22    0.8294
## Year2009           0.22312    0.36600    0.61    0.5440
## Year2010           0.19675    0.41195    0.48    0.6344
## Year2011          -0.27897    0.34834   -0.80    0.4258
## Year2012          -0.00808    0.40683   -0.02    0.9842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.152, Adjusted R-squared:  -0.0482
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 82 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.673  0.927  0.961  0.939  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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-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 1.747 1          1.322
## Year              1.747 16          1.018

```

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
## -1.2306 -0.3698 -0.0285 0.3909 1.0288
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06150 0.40202 2.64 0.01 *
## LastAuthorFemale1 -0.09431 0.19417 -0.49 0.63
## Year1997 -0.19839 0.43980 -0.45 0.65
## Year1998 0.17699 0.40970 0.43 0.67
## Year1999 0.28365 0.48320 0.59 0.56
## Year2000 0.00266 0.41908 0.01 0.99
## Year2001 -0.11660 0.44402 -0.26 0.79
## Year2002 -0.26819 0.84159 -0.32 0.75
## Year2003 -0.30831 0.43665 -0.71 0.48
## Year2004 -0.37005 0.47020 -0.79 0.43
## Year2005 -0.17239 0.46092 -0.37 0.71
## Year2006 0.16911 0.43869 0.39 0.70
```

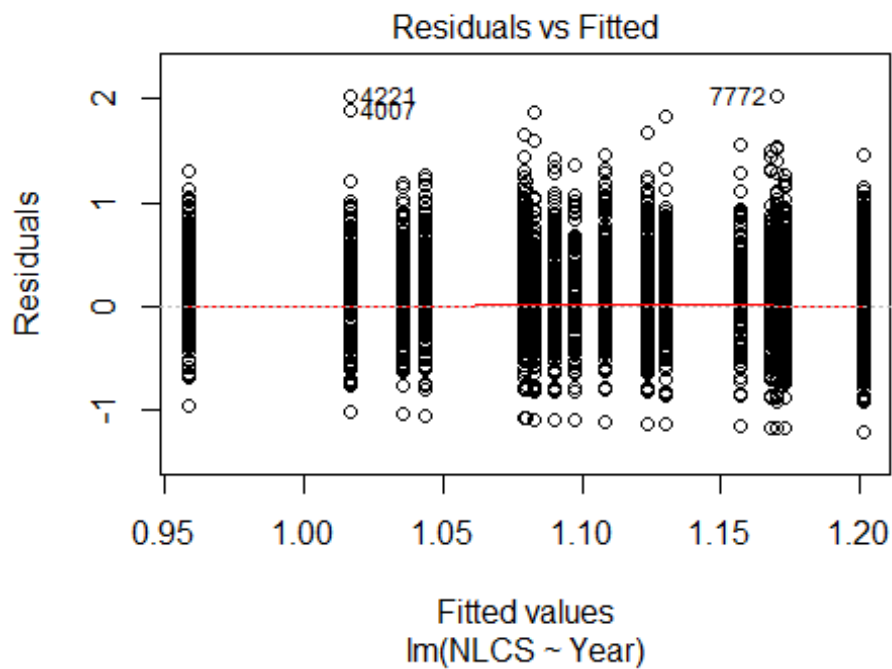


```

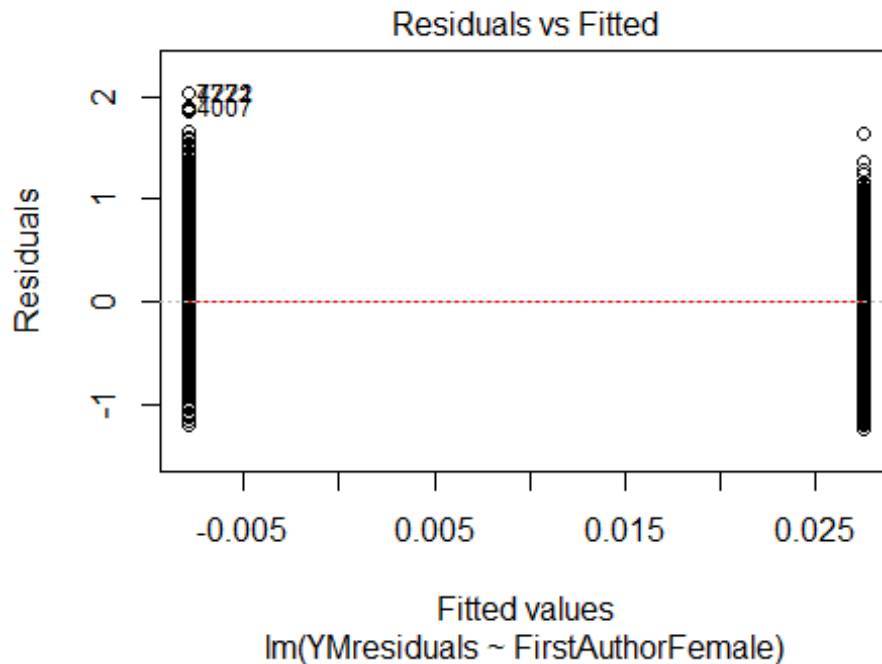
## Year2007          -0.05568      0.42914    -0.13      0.90
## Year2008          -0.09729      0.46564    -0.21      0.84
## Year2009           0.18793      0.47126     0.40      0.69
## Year2010           0.16610      0.45780     0.36      0.72
## Year2011          -0.28934      0.44898    -0.64      0.52
## Year2012          -0.03171      0.46921    -0.07      0.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.138, Adjusted R-squared:  -0.0651
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 82 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.610  0.925  0.954  0.934  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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-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: 90"
## [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
##  516  470  439  409  485  447  428  414  426  403  411  457  465  647  666
## 2011 2012
##  727  780
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  305  280  272  250  252  146  277  235  268  244  251  287  267  377  377
## 2011 2012

```

```
## 423 442
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 290 252 241 225 224 134 248 209 242 212 223 254 231 323 312
## 2011 2012
## 346 363
## [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
```

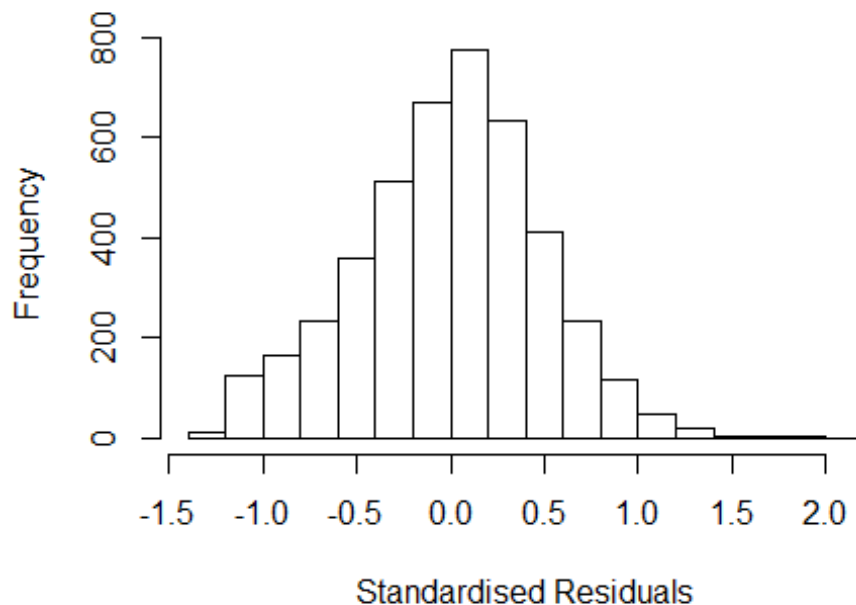


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 4e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.55125927330411"
## [1] "Male first author team size 2018 geometric mean: 3.7190465107968"
##
## 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: 3.95850960548373"
## [1] "Male last author team size 2018 geometric mean: 3.94040074769931"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8900, 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 1.044 1          1.022
## LastAuthorFemale  1.033 1          1.016
## UniqueAuthors    1.154 4          1.018
## Year              1.188 16         1.005
```

Residuals from first and last author and team size



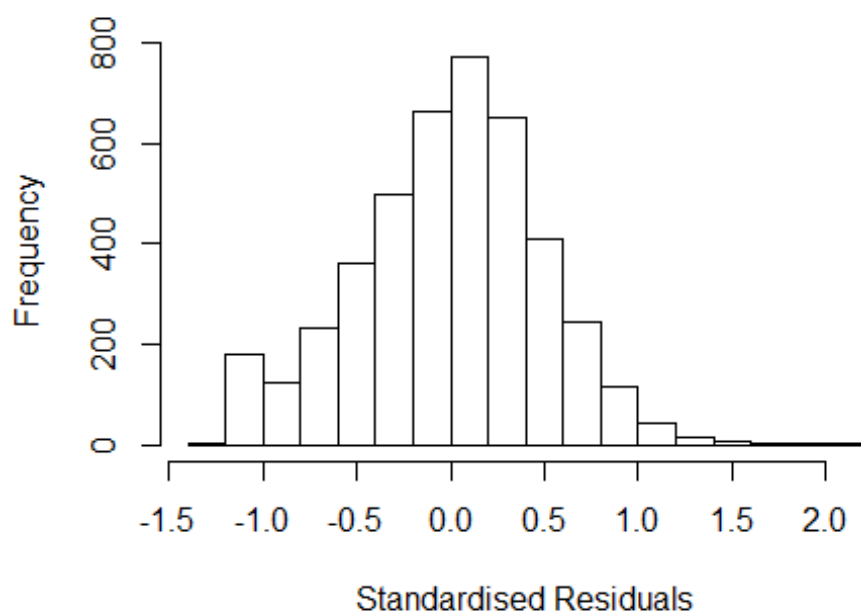
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2659 -0.3148 0.0261 0.3160 2.0330
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.989631 0.047689 20.75 < 2e-16 ***
## FirstAuthorFemale1 0.024248 0.017969 1.35 0.17727
## LastAuthorFemale1 0.028617 0.022651 1.26 0.20651
## UniqueAuthors2 0.133915 0.038976 3.44 0.00060 ***
## UniqueAuthors3 0.057806 0.036739 1.57 0.11569
## UniqueAuthors4 0.122515 0.037010 3.31 0.00094 ***
## UniqueAuthors5 0.213642 0.036238 5.90 4e-09 ***
## Year1997 0.044811 0.044105 1.02 0.30968
## Year1998 0.048794 0.041680 1.17 0.24180
## Year1999 -0.012750 0.045162 -0.28 0.77771
```

```

## Year2000      -0.020679    0.044544   -0.46   0.64249
## Year2001      -0.007120    0.052079   -0.14   0.89126
## Year2002       0.000381    0.043467    0.01   0.99301
## Year2003      -0.119504    0.046570   -2.57   0.01032 *
## Year2004      -0.160092    0.048085   -3.33   0.00088 ***
## Year2005      -0.054344    0.046959   -1.16   0.24723
## Year2006      -0.050915    0.049535   -1.03   0.30407
## Year2007      -0.051715    0.049186   -1.05   0.29313
## Year2008      -0.031317    0.046524   -0.67   0.50089
## Year2009       0.062650    0.043042    1.46   0.14559
## Year2010       0.056700    0.041071    1.38   0.16749
## Year2011       0.074736    0.040155    1.86   0.06278 .
## Year2012      -0.015754    0.040469   -0.39   0.69708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0379, Adjusted R-squared:  0.033
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 351 weights are ~= 1. The remaining 3978 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0206 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          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 1.034 1          1.017
## LastAuthorFemale 1.025 1          1.012
## Year              1.051 16          1.002

```

Residuals from first and last author



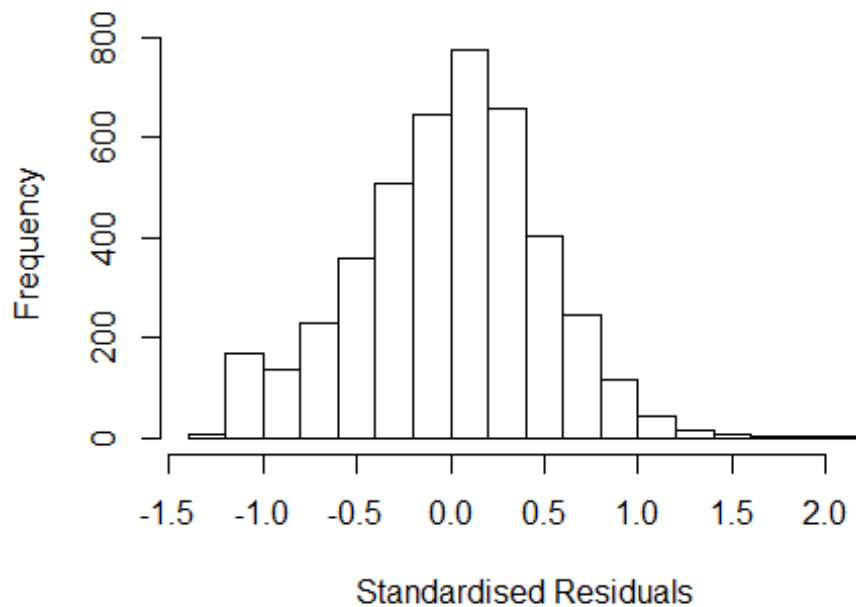
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2361 -0.3241  0.0275  0.3168  2.0572
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10889    0.03195   34.71  < 2e-16 ***
## FirstAuthorFemale1 0.03634    0.01799    2.02  0.04340 *
## LastAuthorFemale1 0.02307    0.02283    1.01  0.31222
## Year1997         0.04120    0.04447    0.93  0.35430
## Year1998         0.04116    0.04235    0.97  0.33113
## Year1999        -0.02373    0.04558   -0.52  0.60265
## Year2000        -0.02493    0.04480   -0.56  0.57793
## Year2001        -0.01858    0.05211   -0.36  0.72151
## Year2002         0.00890    0.04414    0.20  0.84016
## Year2003        -0.12911    0.04695   -2.75  0.00599 **
## Year2004        -0.16721    0.04843   -3.45  0.00056 ***
## Year2005        -0.06681    0.04715   -1.42  0.15656
```

```

## Year2006      -0.06239    0.04928   -1.27  0.20563
## Year2007      -0.05156    0.04952   -1.04  0.29782
## Year2008      -0.01649    0.04698   -0.35  0.72554
## Year2009       0.06778    0.04368    1.55  0.12079
## Year2010       0.06947    0.04161    1.67  0.09506 .
## Year2011       0.08717    0.04063    2.15  0.03197 *
## Year2012       0.00534    0.04097    0.13  0.89623
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0198, Adjusted R-squared:  0.0157
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 370 weights are ~= 1. The remaining 3959 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0191 0.8600 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      2.31e-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 1.029 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
## -1.2193 -0.3250  0.0272  0.3153  2.0549
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11107    0.03179   34.95  <2e-16 ***
## FirstAuthorFemale1 0.03757    0.01795    2.09  0.0365 *
## Year1997        0.04087    0.04442    0.92  0.3576
## Year1998        0.04122    0.04231    0.97  0.3300
## Year1999       -0.02294    0.04556   -0.50  0.6146
## Year2000       -0.02484    0.04480   -0.55  0.5794
## Year2001       -0.01777    0.05212   -0.34  0.7332
## Year2002        0.00931    0.04414    0.21  0.8330
## Year2003       -0.12899    0.04695   -2.75  0.0060 **
## Year2004       -0.16612    0.04839   -3.43  0.0006 ***
## Year2005       -0.06594    0.04716   -1.40  0.1621
## Year2006       -0.06185    0.04933   -1.25  0.2100
```



```

## Year2007          -0.05141    0.04954   -1.04    0.2995
## Year2008          -0.01638    0.04691   -0.35    0.7270
## Year2009           0.06769    0.04365    1.55    0.1211
## Year2010           0.07069    0.04160    1.70    0.0893 .
## Year2011           0.08916    0.04058    2.20    0.0281 *
## Year2012           0.00707    0.04091    0.17    0.8629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0195, Adjusted R-squared:  0.0156
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 369 weights are ~= 1. The remaining 3960 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0195 0.8600 0.9500 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      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 1.021 1      1.010
## 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
## -1.2321 -0.3257  0.0297  0.3176  2.0473

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11453    0.03189   34.95 < 2e-16 ***
## LastAuthorFemale1 0.02602    0.02279    1.14  0.25363
## Year1997        0.04189    0.04437    0.94  0.34520
## Year1998        0.04219    0.04248    0.99  0.32071
## Year1999       -0.02187    0.04560   -0.48  0.63160
## Year2000       -0.02283    0.04487   -0.51  0.61093
## Year2001       -0.01953    0.05227   -0.37  0.70867
## Year2002        0.01054    0.04405    0.24  0.81096
## Year2003       -0.12486    0.04700   -2.66  0.00792 **
## Year2004       -0.16554    0.04849   -3.41  0.00065 ***
## Year2005       -0.06450    0.04723   -1.37  0.17208
## Year2006       -0.06183    0.04933   -1.25  0.21010
## Year2007       -0.04958    0.04956   -1.00  0.31713
## Year2008       -0.01319    0.04713   -0.28  0.77951
## Year2009        0.07218    0.04355    1.66  0.09751 .
## Year2010        0.07317    0.04163    1.76  0.07886 .
## Year2011        0.09151    0.04058    2.25  0.02420 *
## Year2012        0.00867    0.04092    0.21  0.83214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0189, Adjusted R-squared:  0.015
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 381 weights are ~= 1. The remaining 3948 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0214 0.8600 0.9500 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      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: 4329"
## [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
##    2   10   27    8   14   11    4   10   16   21   22   23   24   16   12
## 2011 2012
##   17   12
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    5   25    6    7    8    3    4   15   19   13   16   17   13   10
## 2011 2012
##   14    6
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    5   25    4    5    6    2    4   14   19   13   16   15   13    8
## 2011 2012
##   12    5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.76644715017318"
## [1] "Male first author team size 2018 geometric mean: 4.91419788067991"

## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.08681028303711"
## [1] "Male last author team size 2018 geometric mean: 4.24687933286967"

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

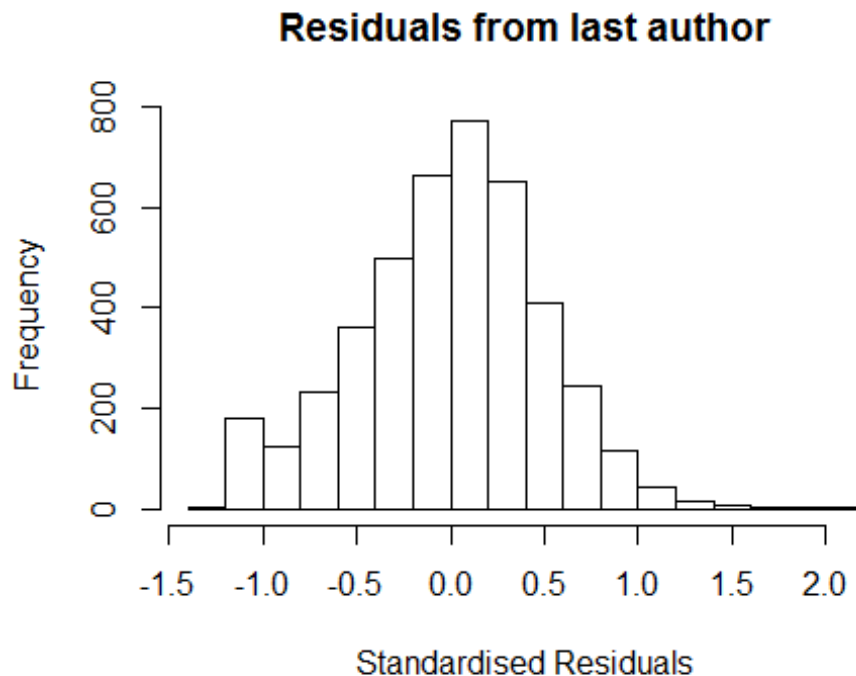
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 56, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

```

```
## Warning in 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"

## 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 1.609  1      1.268
## LastAuthorFemale  2.288  1      1.513
## Year              3.533 16      1.040

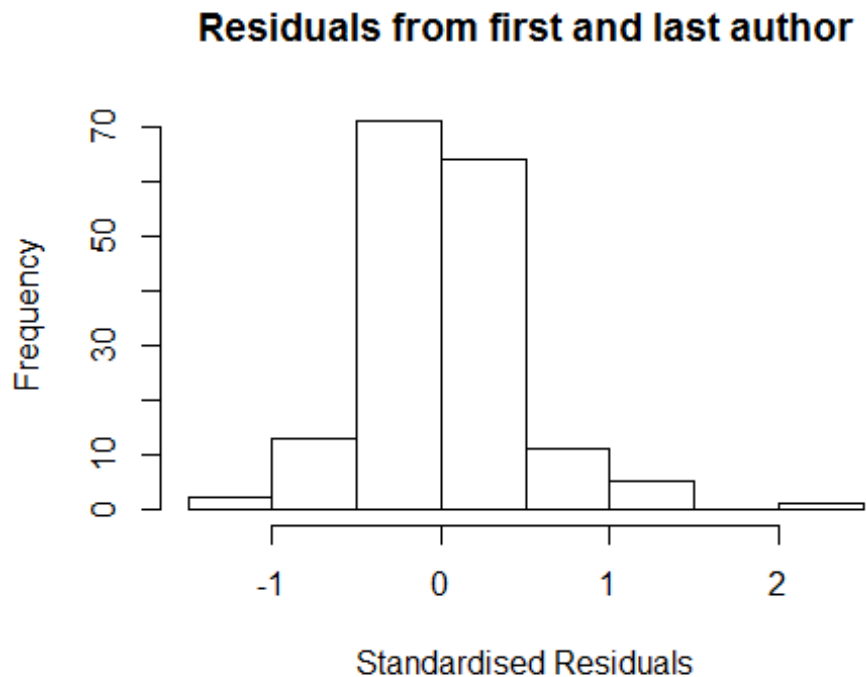
## [1] "List of  0 outliers with residuals above  2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2525 -0.2007 -0.0314  0.2469  2.0110
##
```

```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.74600    0.00000      Inf < 2e-16 ***
## FirstAuthorFemale1 0.00596    0.08539     0.07 0.94449
## LastAuthorFemale1 0.26332    0.09042     2.91 0.00415 **
## Year1997        0.08201    0.58738     0.14 0.88914
## Year1998       -0.64584    0.03618   -17.85 < 2e-16 ***
## Year1999        0.33847    0.14515     2.33 0.02106 *
## Year2000       -0.30504    0.30357    -1.00 0.31661
## Year2001        0.70124    0.30572     2.29 0.02321 *
## Year2002        0.04084    0.14151     0.29 0.77329
## Year2003        0.63386    0.15401     4.12 6.4e-05 ***
## Year2004        0.30473    0.17316     1.76 0.08051 .
## Year2005        0.08681    0.10950     0.79 0.42917
## Year2006        0.45295    0.11423     3.97 0.00011 ***
## Year2007        0.38372    0.08658     4.43 1.8e-05 ***
## Year2008        0.43643    0.10465     4.17 5.2e-05 ***
## Year2009        0.08056    0.14397     0.56 0.57660
## Year2010        0.23720    0.34151     0.69 0.48841
## Year2011        0.25341    0.07079     3.58 0.00047 ***
## Year2012        0.00573    0.21002     0.03 0.97828
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.549, Adjusted R-squared:  0.494
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.0006);
## 7 weights are ~= 1. The remaining 159 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0567 0.8800 0.9650 0.8880 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           5.99e-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"

```

```
## 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 1.456 1          1.207
## Year              1.456 16         1.012

## [1] "List of 0 outliers 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
## -1.0711 -0.2342 -0.0143  0.2557  1.9735
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7460    0.0000      Inf < 2e-16 ***
## FirstAuthorFemale1 0.0443    0.0904    0.49  0.6248
## Year1997        0.1195    0.6336    0.19  0.8506
## Year1998       -0.6452    0.0361   -17.86 < 2e-16 ***
## Year1999        0.3279    0.1478    2.22  0.0281 *
```

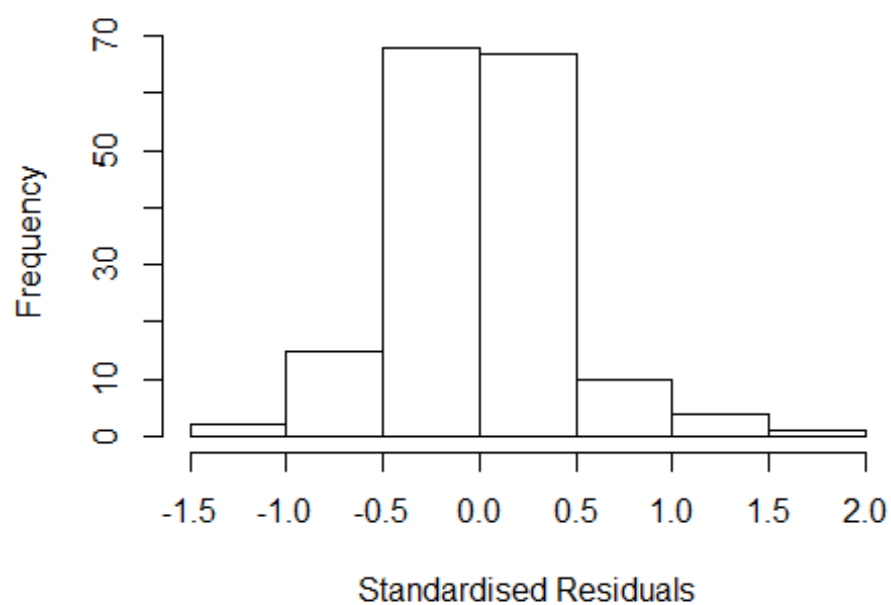
```

## Year2000          -0.2087      0.3701    -0.56    0.5737
## Year2001          0.6993      0.3042      2.30    0.0229 *
## Year2002          0.1725      0.2529      0.68    0.4962
## Year2003          0.6206      0.1520      4.08    7.2e-05 ***
## Year2004          0.3455      0.1664      2.08    0.0396 *
## Year2005          0.1760      0.1033      1.70    0.0905 .
## Year2006          0.4906      0.1094      4.48    1.5e-05 ***
## Year2007          0.4417      0.0840      5.26    4.9e-07 ***
## Year2008          0.4626      0.1052      4.40    2.1e-05 ***
## Year2009          0.1198      0.1444      0.83    0.4084
## Year2010          0.2808      0.3425      0.82    0.4136
## Year2011          0.2750      0.0827      3.32    0.0011 **
## Year2012          0.1180      0.1811      0.65    0.5157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.507, Adjusted R-squared:  0.45
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.0006);
## 11 weights are ~= 1. The remaining 155 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.125  0.879  0.957   0.892  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           5.99e-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"

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

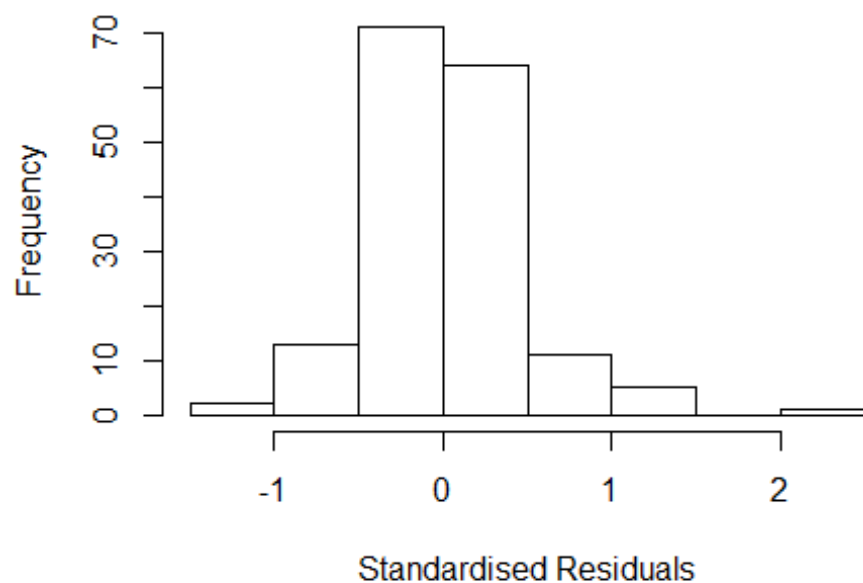
```

Residuals from first author



##	GVIF	Df	$GVIF^{(1/(2*Df))}$
## LastAuthorFemale	2.253	1	1.501
## Year	2.253	16	1.026

Residuals from last author

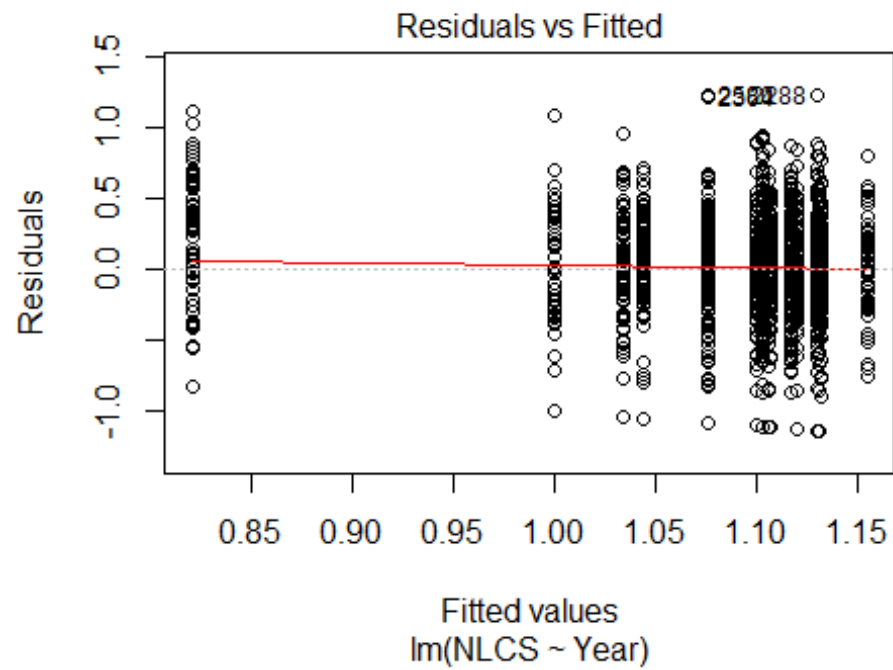



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2396 -0.2033 -0.0326 0.2476 2.0005
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.74600 0.00000 Inf < 2e-16 ***
## LastAuthorFemale1 0.26231 0.09286 2.82 0.00538 **
## Year1997 0.09246 0.60758 0.15 0.87925
## Year1998 -0.64563 0.03616 -17.86 < 2e-16 ***
## Year1999 0.34010 0.14076 2.42 0.01690 *
## Year2000 -0.29949 0.30101 -0.99 0.32138
## Year2001 0.70638 0.29761 2.37 0.01889 *
## Year2002 0.04135 0.14207 0.29 0.77144
## Year2003 0.63480 0.15057 4.22 4.3e-05 ***
## Year2004 0.30956 0.17160 1.80 0.07326 .
## Year2005 0.08963 0.10851 0.83 0.41013
## Year2006 0.45354 0.11523 3.94 0.00013 ***
## Year2007 0.38544 0.08581 4.49 1.4e-05 ***
## Year2008 0.43767 0.09789 4.47 1.5e-05 ***
## Year2009 0.08307 0.14140 0.59 0.55777
## Year2010 0.23128 0.33830 0.68 0.49526
## Year2011 0.25459 0.06588 3.86 0.00017 ***
## Year2012 0.00753 0.20456 0.04 0.97067
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared: 0.544, Adjusted R-squared: 0.492
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 0.0006);
## 6 weights are ~ = 1. The remaining 160 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0759 0.8840 0.9670 0.8920 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 5.99e-04 1.82e-12
```

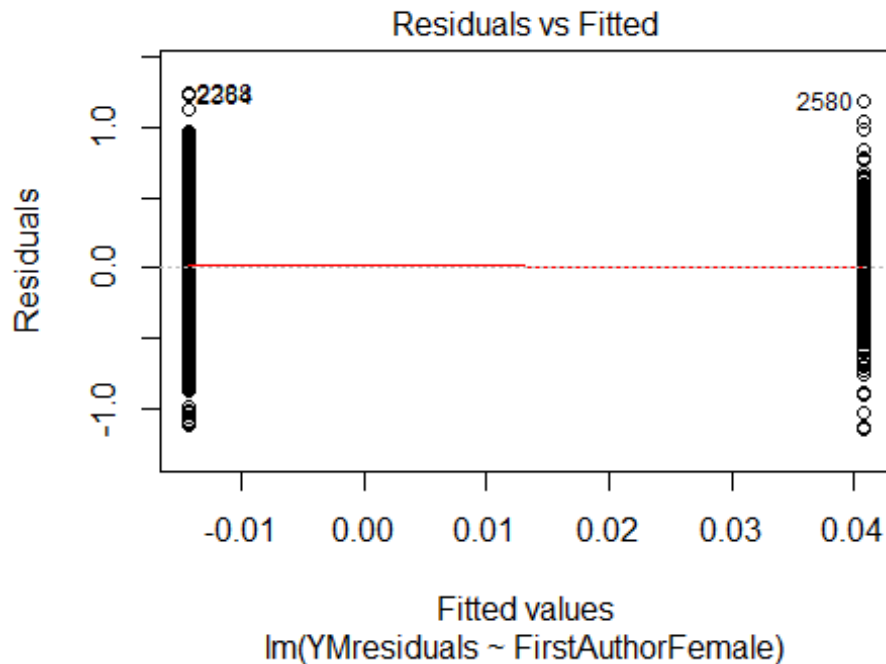
```

## 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: 167"
## [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
## 138 140 134 112 102 110 103 83 105 124 132 130 159 165 183
## 2011 2012
## 216 200
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 78 89 60 62 52 70 48 76 90 91 82 99 114 119
## 2011 2012
## 150 139
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 64 77 54 51 46 60 37 65 83 77 73 89 101 103
## 2011 2012
## 128 110
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 78, df = 16, p-value = 4e-10

```

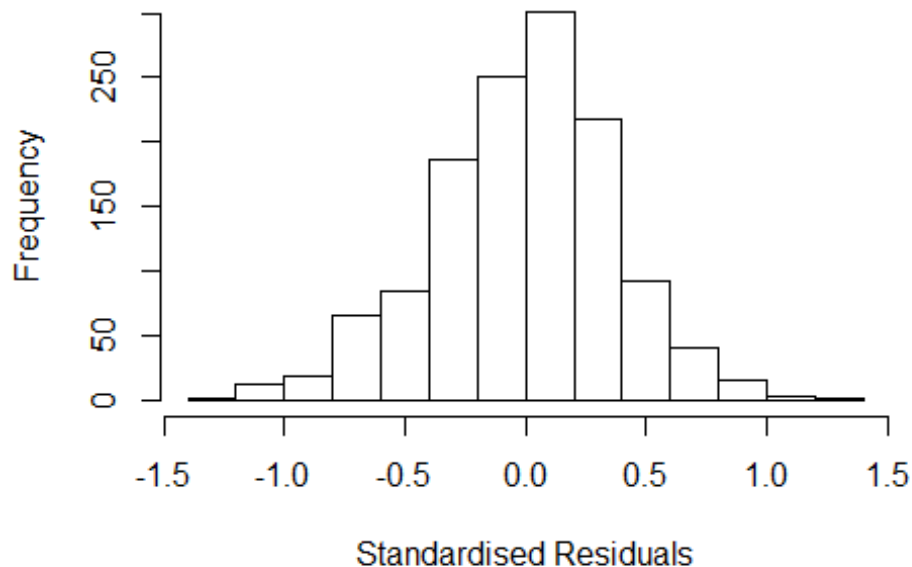


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.9, df = 1, p-value = 0.09
```



```
## [1] "Female first author team size 2018 geometric mean: 4.78932594032853"
## [1] "Male first author team size 2018 geometric mean: 4.04741249263074"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.42432317336591"
## [1] "Male last author team size 2018 geometric mean: 4.32042502404439"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 570, 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 1.088 1      1.043
## LastAuthorFemale  1.109 1      1.053
## UniqueAuthors    1.589 4      1.060
## Year              1.733 16     1.017
```

Residuals from first and last author and team size



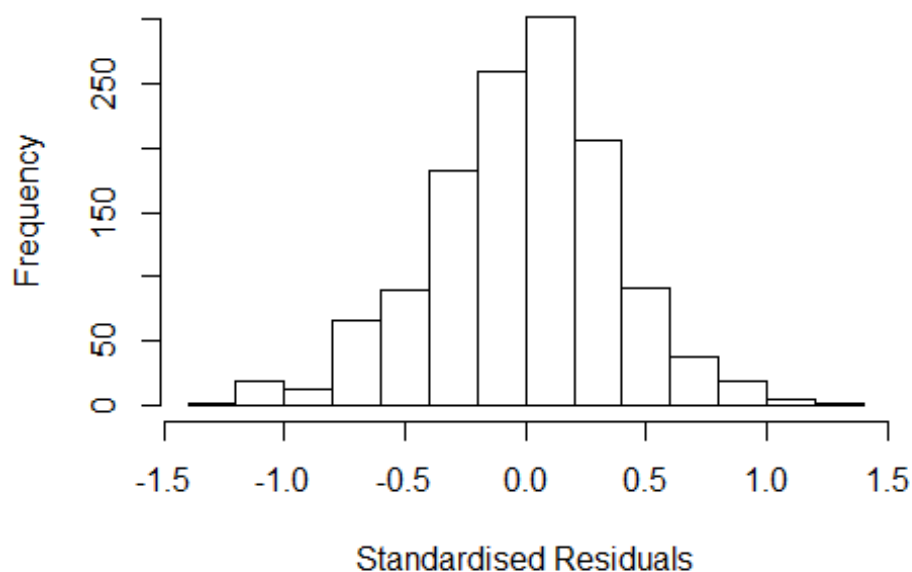
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2248 -0.2484  0.0199  0.2290  1.2322
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08567    0.07196   15.09  <2e-16 ***
## FirstAuthorFemale1  0.01632    0.02321    0.70   0.4822
## LastAuthorFemale1  0.04251    0.03012    1.41   0.1584
## UniqueAuthors2   -0.02746    0.06474   -0.42   0.6716
## UniqueAuthors3    0.09032    0.06251    1.44   0.1488
## UniqueAuthors4    0.11439    0.06375    1.79   0.0730 .
## UniqueAuthors5    0.14395    0.06268    2.30   0.0218 *
## Year1997         -0.11055    0.07137   -1.55   0.1216
## Year1998         -0.35513    0.12367   -2.87   0.0042 **
## Year1999         -0.00488    0.07364   -0.07   0.9472
```

```

## Year2000      0.00957    0.07370    0.13    0.8967
## Year2001     -0.15542    0.08546   -1.82    0.0692 .
## Year2002     -0.05481    0.06987   -0.78    0.4329
## Year2003     -0.00943    0.07658   -0.12    0.9021
## Year2004     -0.04292    0.05906   -0.73    0.4675
## Year2005     -0.06053    0.05695   -1.06    0.2880
## Year2006     -0.14569    0.05933   -2.46    0.0142 *
## Year2007     -0.07236    0.05976   -1.21    0.2262
## Year2008     -0.06611    0.05715   -1.16    0.2476
## Year2009     -0.09614    0.05763   -1.67    0.0955 .
## Year2010     -0.07423    0.05801   -1.28    0.2009
## Year2011     -0.06612    0.05638   -1.17    0.2412
## Year2012      0.00534    0.06132    0.09    0.9306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared:  0.0851, Adjusted R-squared:  0.0692
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.864  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      7.76e-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 1.061 1      1.030
## LastAuthorFemale  1.081 1      1.040
## Year              1.137 16      1.004

```

Residuals from first and last author



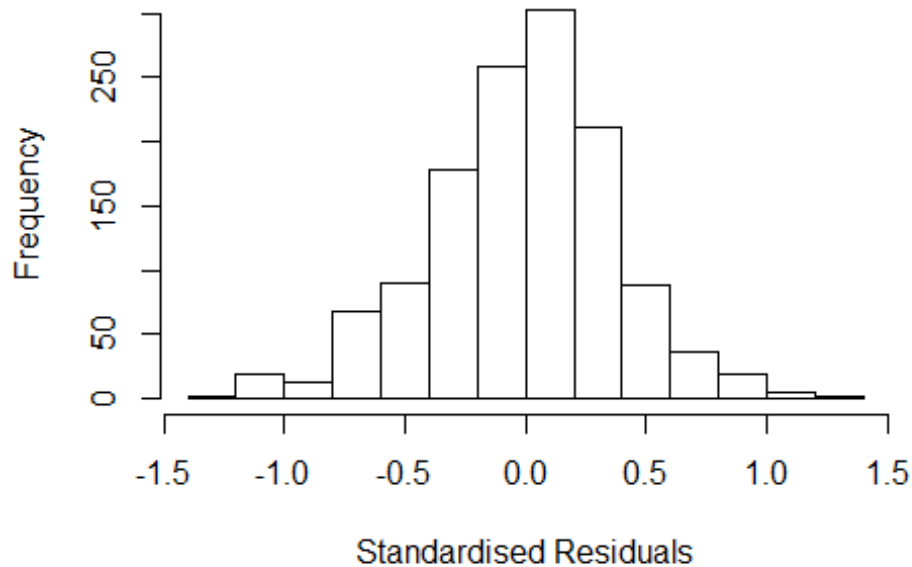
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2179 -0.2395  0.0172  0.2320  1.2538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.127463   0.046962   24.01  <2e-16 ***
## FirstAuthorFemale1 0.029140   0.023128    1.26   0.208
## LastAuthorFemale1 0.036943   0.029677    1.24   0.213
## Year1997        -0.097954   0.071512   -1.37   0.171
## Year1998        -0.362723   0.141731   -2.56   0.011 *
## Year1999         0.003372   0.073274    0.05   0.963
## Year2000         0.015728   0.072946    0.22   0.829
## Year2001        -0.113967   0.088022   -1.29   0.196
## Year2002        -0.012674   0.068003   -0.19   0.852
## Year2003         0.028179   0.075910    0.37   0.711
## Year2004         0.000732   0.056849    0.01   0.990
## Year2005        -0.005214   0.055733   -0.09   0.925
```

```

## Year2006      -0.100981    0.058730   -1.72    0.086 .
## Year2007      -0.022884    0.059721   -0.38    0.702
## Year2008      -0.036159    0.057192   -0.63    0.527
## Year2009      -0.047625    0.057353   -0.83    0.406
## Year2010      -0.023282    0.057697   -0.40    0.687
## Year2011      -0.023372    0.056530   -0.41    0.679
## Year2012       0.061289    0.060418    1.01    0.311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.346
## Multiple R-squared:  0.0565, Adjusted R-squared:  0.0431
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1192 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.160  0.866  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      7.76e-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 1.056 1      1.028
## Year              1.056 16      1.002

```


Residuals from first author



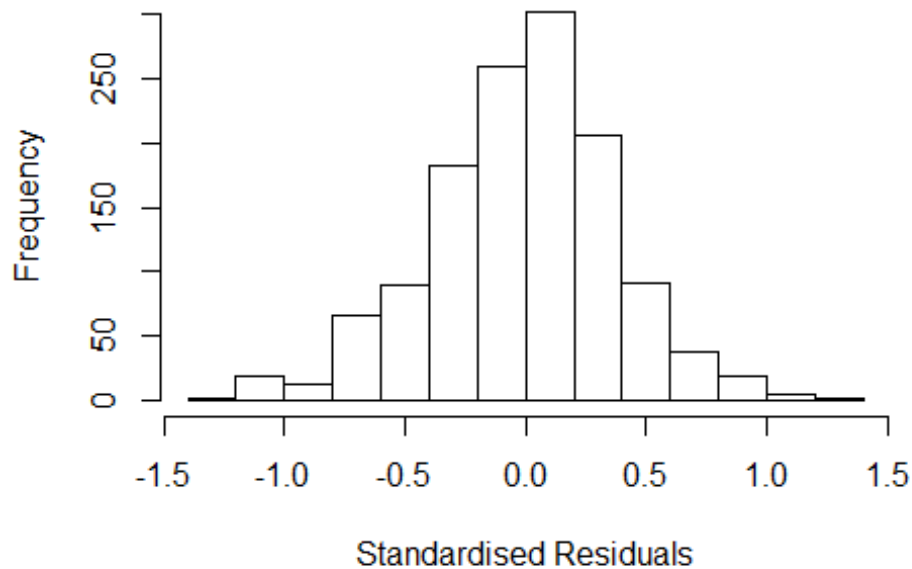
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2243 -0.2416  0.0201  0.2292  1.2466
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13015    0.04684   24.13  <2e-16 ***
## FirstAuthorFemale1 0.02878    0.02319    1.24   0.215
## Year1997      -0.09843    0.07149   -1.38   0.169
## Year1998      -0.36282    0.14500   -2.50   0.012 *
## Year1999       0.00540    0.07317    0.07   0.941
## Year2000       0.01551    0.07341    0.21   0.833
## Year2001      -0.11299    0.08825   -1.28   0.201
## Year2002      -0.01242    0.06795   -0.18   0.855
## Year2003       0.02771    0.07625    0.36   0.716
## Year2004       0.00258    0.05664    0.05   0.964
## Year2005      -0.00111    0.05542   -0.02   0.984
## Year2006      -0.09889    0.05874   -1.68   0.093 .
```

```

## Year2007          -0.02074      0.05953    -0.35      0.728
## Year2008          -0.03133      0.05676    -0.55      0.581
## Year2009          -0.04200      0.05697    -0.74      0.461
## Year2010          -0.01878      0.05744    -0.33      0.744
## Year2011          -0.01607      0.05595    -0.29      0.774
## Year2012           0.06541      0.06003      1.09      0.276
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.344
## Multiple R-squared:  0.0554, Adjusted R-squared:  0.0427
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1184 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.160  0.863  0.949  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      7.76e-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 1.077 1          1.038
## Year              1.077 16          1.002

```

Residuals from last author



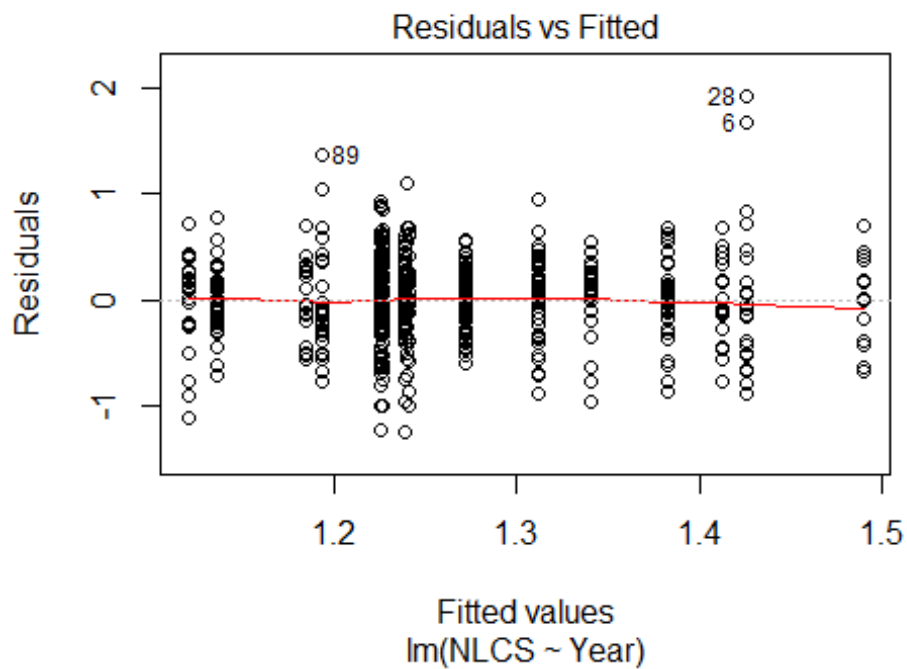
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1980 -0.2393 0.0182 0.2276 1.2468
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.134971 0.046469 24.42 <2e-16 ***
## LastAuthorFemale1 0.036461 0.029819 1.22 0.222
## Year1997 -0.098373 0.071754 -1.37 0.171
## Year1998 -0.364119 0.147314 -2.47 0.014 *
## Year1999 0.002471 0.072906 0.03 0.973
## Year2000 0.016028 0.073397 0.22 0.827
## Year2001 -0.112830 0.087784 -1.29 0.199
## Year2002 -0.015724 0.068002 -0.23 0.817
## Year2003 0.027397 0.075813 0.36 0.718
## Year2004 -0.000173 0.056725 0.00 0.998
## Year2005 -0.002628 0.055775 -0.05 0.962
## Year2006 -0.099181 0.058731 -1.69 0.092 .
```

```

## Year2007          -0.020511    0.059410    -0.35    0.730
## Year2008          -0.035280    0.056910    -0.62    0.535
## Year2009          -0.044700    0.057191    -0.78    0.435
## Year2010          -0.023755    0.057701    -0.41    0.681
## Year2011          -0.022178    0.056468    -0.39    0.695
## Year2012           0.062999    0.060481     1.04    0.298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared:  0.055, Adjusted R-squared:  0.0424
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1182 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.160  0.862  0.949  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      7.76e-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: 1289"
## [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
##   52   65   37   45   55   63   51   40   41   47   73   71  105   95  116
## 2011 2012
##  113  103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   27   17   18   24   18   28   18   13   22   36   32   54   59   64
## 2011 2012

```

```
## 69 61
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 23 20 16 16 19 17 25 14 11 17 26 22 46 48 54
## 2011 2012
## 56 52
## [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 = 9e-07
```



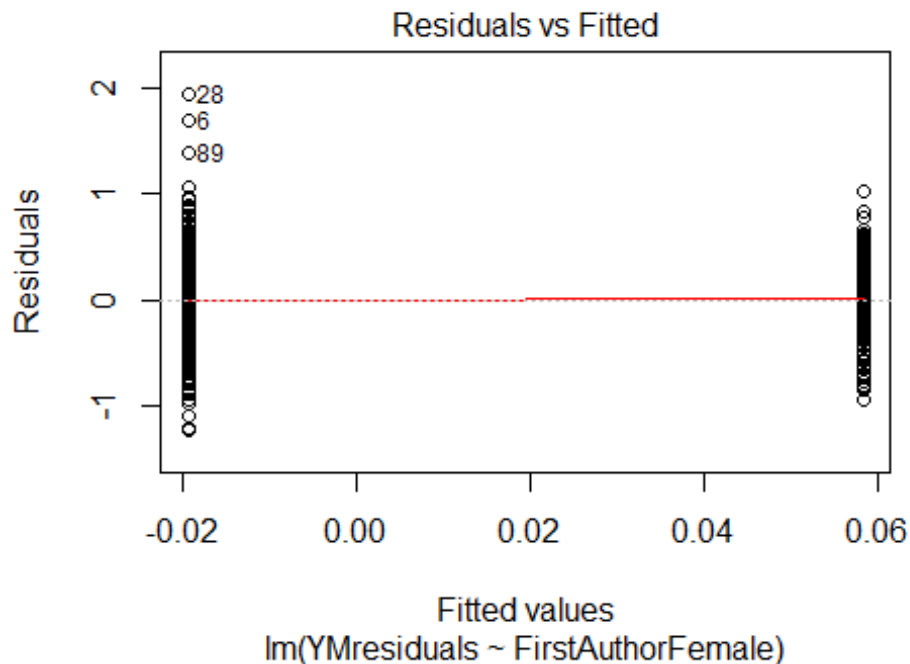
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 4.92380085365595"
## [1] "Male first author team size 2018 geometric mean: 4.52915619705883"

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

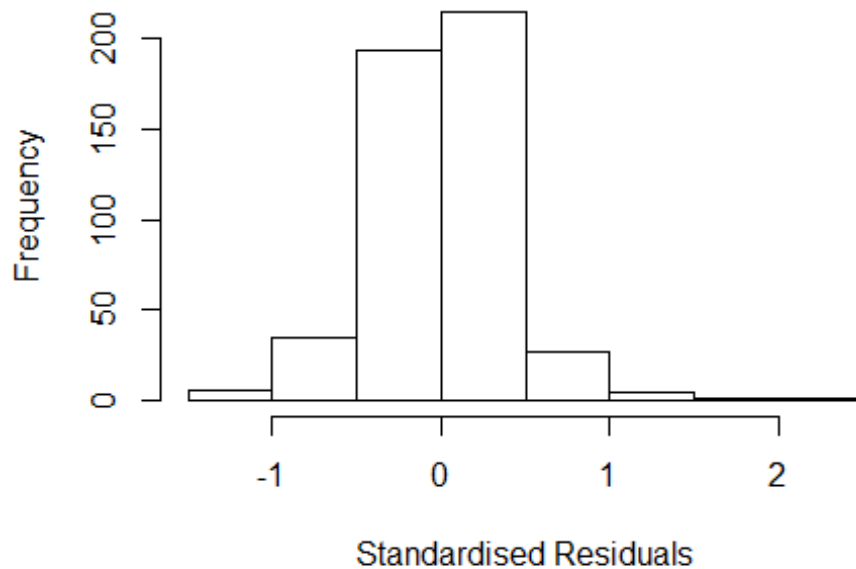
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 78, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.0356543087298"
## [1] "Male last author team size 2018 geometric mean: 4.88681812404464"

## 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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.225  1      1.107
## LastAuthorFemale  1.393  1      1.180
## UniqueAuthors    2.971  4      1.146
## Year              3.558 16      1.040
```

Residuals from first and last author and team size



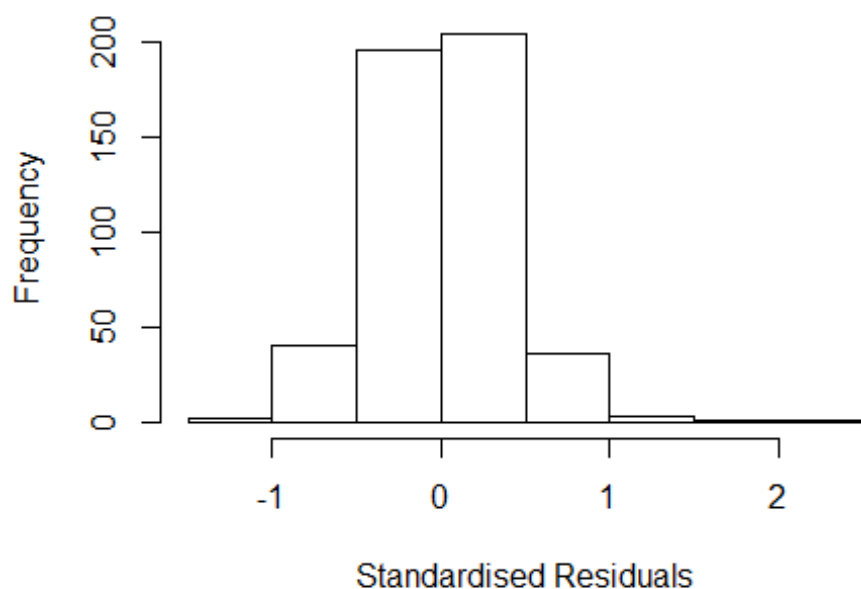
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32605 -0.23086 0.00527 0.23082 2.22354
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12046 0.15517 7.22 2.2e-12 ***
## FirstAuthorFemale1 0.06438 0.04124 1.56 0.119
## LastAuthorFemale1 -0.02050 0.05255 -0.39 0.697
## UniqueAuthors2 0.01031 0.09916 0.10 0.917
## UniqueAuthors3 0.04394 0.09355 0.47 0.639
## UniqueAuthors4 0.15767 0.09463 1.67 0.096 .
## UniqueAuthors5 0.20423 0.09104 2.24 0.025 *
## Year1997 -0.00784 0.17734 -0.04 0.965
## Year1998 0.03194 0.17751 0.18 0.857
## Year1999 0.21887 0.15299 1.43 0.153
```

```

## Year2000      0.17431    0.14061    1.24    0.216
## Year2001      0.17736    0.18133    0.98    0.329
## Year2002      0.00259    0.13369    0.02    0.985
## Year2003      0.05201    0.14960    0.35    0.728
## Year2004      0.31379    0.21439    1.46    0.144
## Year2005     -0.04430    0.15581   -0.28    0.776
## Year2006     -0.03187    0.12984   -0.25    0.806
## Year2007      0.12364    0.13864    0.89    0.373
## Year2008      0.06260    0.12992    0.48    0.630
## Year2009     -0.04739    0.13027   -0.36    0.716
## Year2010      0.00137    0.12829    0.01    0.991
## Year2011      0.03065    0.12399    0.25    0.805
## Year2012     -0.04588    0.12726   -0.36    0.719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.339
## Multiple R-squared:  0.0946, Adjusted R-squared:  0.0512
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(3,12) are outliers with |weight| = 0 ( < 0.00021);
## 53 weights are ~= 1. The remaining 427 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0922 0.8590 0.9480 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      2.07e-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 1.207 1 1.099
## LastAuthorFemale 1.243 1 1.115
## Year 1.358 16 1.010

```


Residuals from first and last author



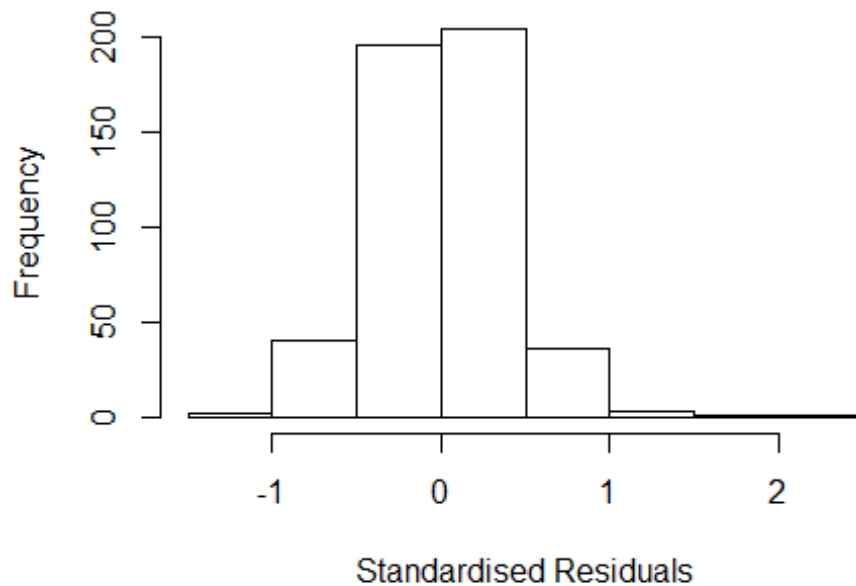
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36573 -0.24137 0.00934 0.22870 2.14673
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19727 0.12234 9.79 <2e-16 ***
## FirstAuthorFemale1 0.04836 0.04182 1.16 0.25
## LastAuthorFemale1 0.01324 0.05026 0.26 0.79
## Year1997 -0.03166 0.17441 -0.18 0.86
## Year1998 0.01738 0.16989 0.10 0.92
## Year1999 0.22674 0.15755 1.44 0.15
## Year2000 0.15958 0.14526 1.10 0.27
## Year2001 0.15522 0.17554 0.88 0.38
## Year2002 -0.00925 0.13784 -0.07 0.95
## Year2003 0.08605 0.14898 0.58 0.56
## Year2004 0.31255 0.20113 1.55 0.12
## Year2005 -0.01438 0.17229 -0.08 0.93
```

```

## Year2006      -0.01332    0.13593   -0.10    0.92
## Year2007      0.19527    0.14257    1.37    0.17
## Year2008      0.09108    0.13591    0.67    0.50
## Year2009     -0.01198    0.13630   -0.09    0.93
## Year2010      0.05318    0.13079    0.41    0.68
## Year2011      0.06550    0.13036    0.50    0.62
## Year2012     -0.01232    0.13404   -0.09    0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0493, Adjusted R-squared:  0.0124
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(3,12) are outliers with |weight| = 0 ( < 0.00021);
## 39 weights are ~= 1. The remaining 441 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.873   0.951   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      2.07e-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 1.152 1      1.073
## Year              1.152 16      1.004

```

Residuals from first author



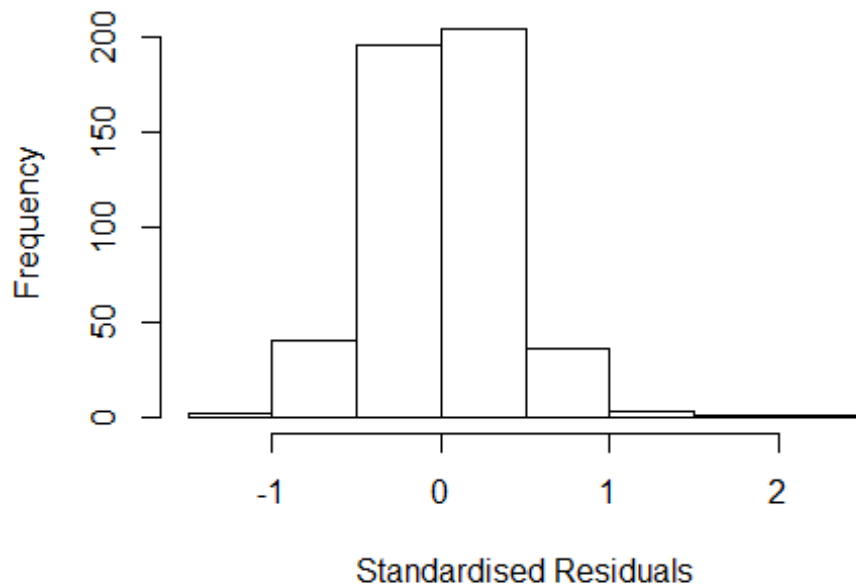
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3516 -0.2432 0.0108 0.2299 2.1472
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19683 0.12255 9.77 <2e-16 ***
## FirstAuthorFemale1 0.04995 0.04095 1.22 0.22
## Year1997 -0.03151 0.17459 -0.18 0.86
## Year1998 0.02100 0.16900 0.12 0.90
## Year1999 0.22828 0.15781 1.45 0.15
## Year2000 0.16135 0.14529 1.11 0.27
## Year2001 0.15475 0.17553 0.88 0.38
## Year2002 -0.00810 0.13819 -0.06 0.95
## Year2003 0.08796 0.14907 0.59 0.56
## Year2004 0.31545 0.20232 1.56 0.12
## Year2005 -0.01158 0.17292 -0.07 0.95
## Year2006 -0.01143 0.13608 -0.08 0.93
```

```

## Year2007          0.19670      0.14303      1.38      0.17
## Year2008          0.09278      0.13606      0.68      0.50
## Year2009         -0.01020      0.13641     -0.07      0.94
## Year2010          0.05595      0.13086      0.43      0.67
## Year2011          0.06752      0.13053      0.52      0.61
## Year2012         -0.00972      0.13387     -0.07      0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared:  0.0493, Adjusted R-squared:  0.0144
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(3,12) are outliers with |weight| = 0 ( < 0.00021);
## 42 weights are ~= 1. The remaining 438 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.110  0.872  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      2.07e-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 1.186 1         1.089
## Year             1.186 16         1.005

```

Residuals from last author



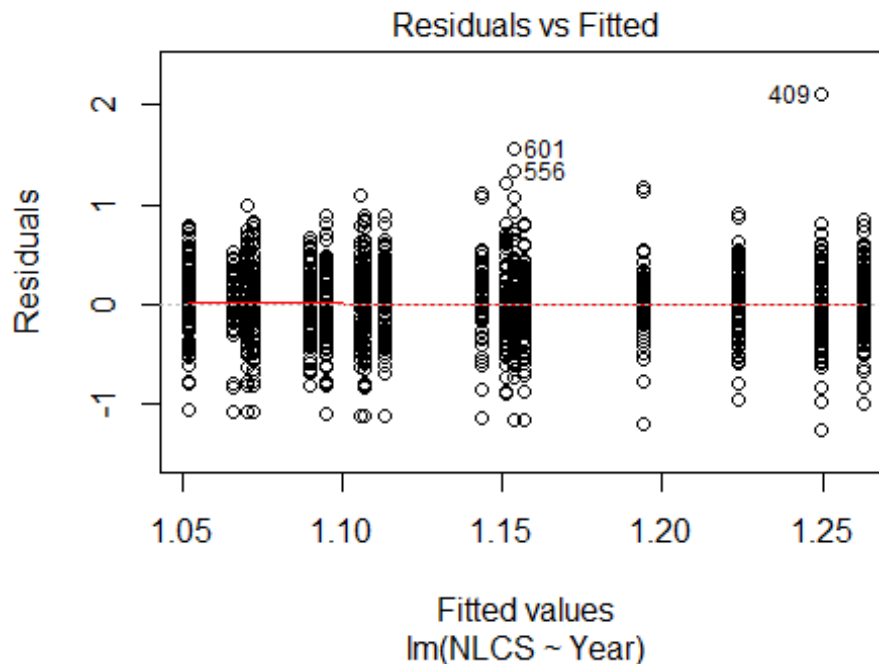
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37973 -0.24583 0.00945 0.22553 2.14220
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.201795 0.119671 10.04 <2e-16 ***
## LastAuthorFemale1 0.022619 0.049229 0.46 0.65
## Year1997 -0.034801 0.173144 -0.20 0.84
## Year1998 0.021223 0.170595 0.12 0.90
## Year1999 0.221357 0.155346 1.42 0.15
## Year2000 0.164385 0.142166 1.16 0.25
## Year2001 0.155314 0.173364 0.90 0.37
## Year2002 -0.005060 0.134687 -0.04 0.97
## Year2003 0.095697 0.146266 0.65 0.51
## Year2004 0.323092 0.198577 1.63 0.10
## Year2005 0.000801 0.167481 0.00 1.00
## Year2006 -0.010430 0.132700 -0.08 0.94
```

```

## Year2007      0.204199    0.140824    1.45    0.15
## Year2008      0.100004    0.132651    0.75    0.45
## Year2009     -0.002108    0.132945   -0.02    0.99
## Year2010      0.061126    0.127234    0.48    0.63
## Year2011      0.072060    0.126796    0.57    0.57
## Year2012     -0.004456    0.130610   -0.03    0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0463, Adjusted R-squared:  0.0114
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(3,12) are outliers with |weight| = 0 ( < 0.00021);
## 44 weights are ~= 1. The remaining 436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0999 0.8730 0.9500 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      2.07e-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: 482"
## [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
## 149 125 160 118 124 131 105 128 136 125 133 153 128 124 136
## 2011 2012
## 147 130
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 82 95 78 65 40 69 79 91 83 107 129 110 92 104

```

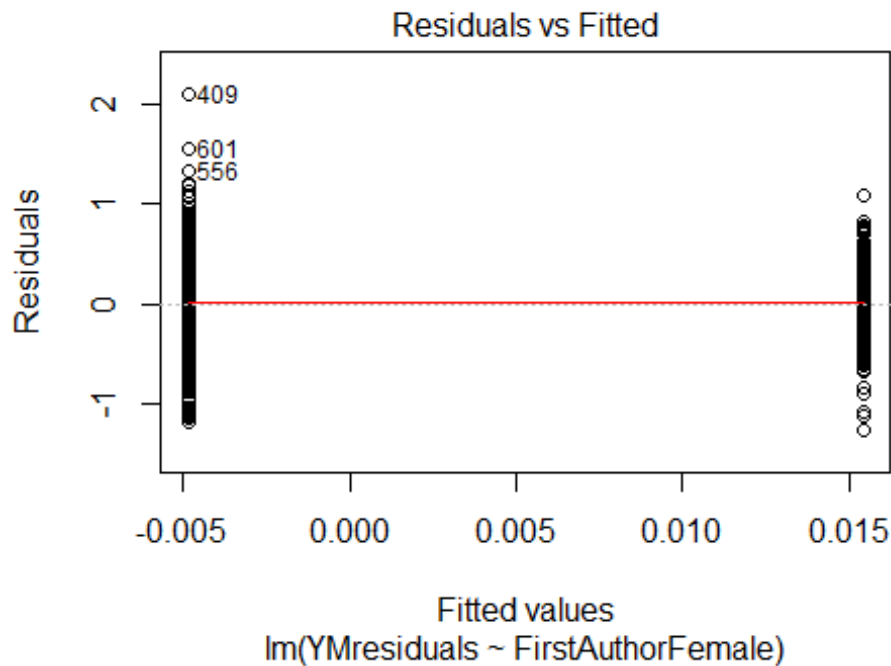
```
## 2011 2012
## 108 95
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 69 83 63 58 32 61 65 83 74 90 110 91 83 95
## 2011 2012
## 94 82
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.81, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 4.24164478453941"
## [1] "Male first author team size 2018 geometric mean: 3.74560822540566"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

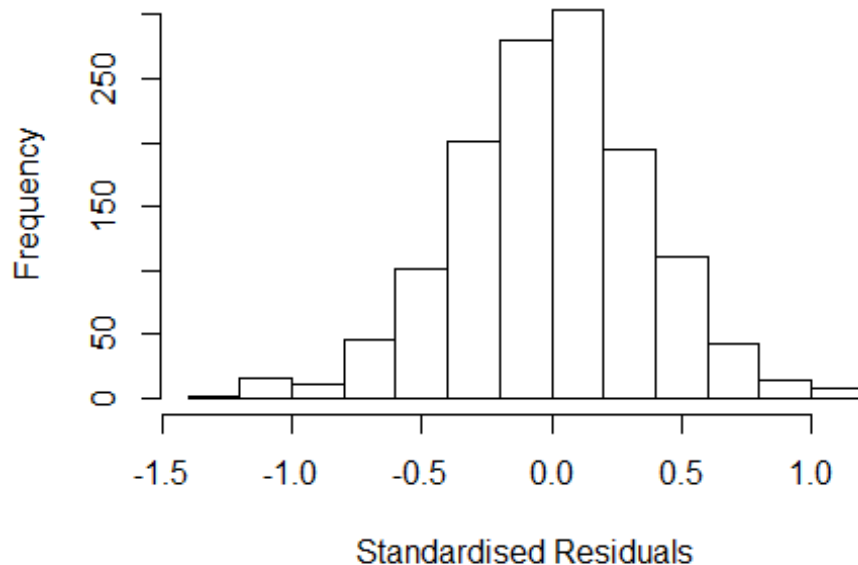
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.30698180344988"
## [1] "Male last author team size 2018 geometric mean: 3.76826653996826"

## 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 = 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 1.074 1      1.036
## LastAuthorFemale  1.040 1      1.020
## UniqueAuthors    1.313 4      1.035
## Year              1.435 16     1.011
```


Residuals from first and last author and team size



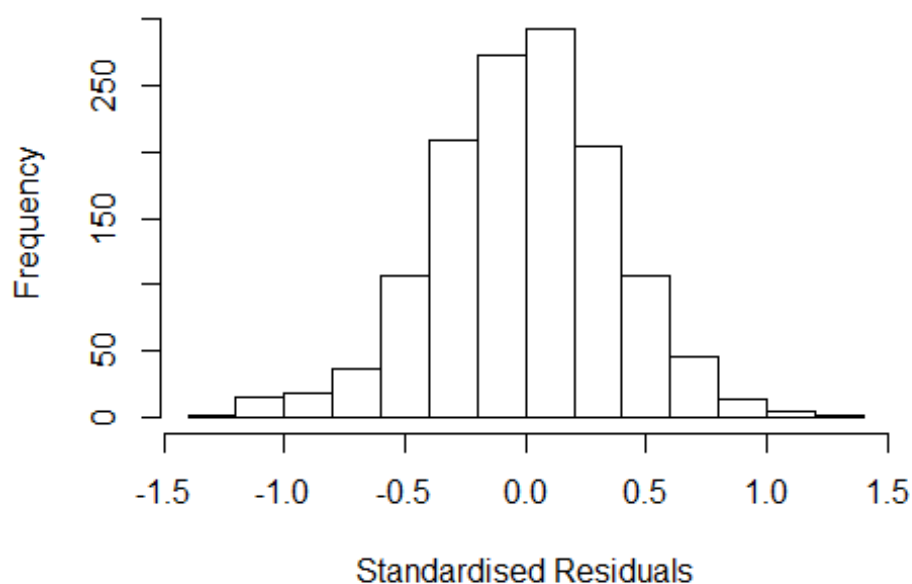
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25048 -0.23966 0.00727 0.22336 1.18253
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3571 0.0822 16.51 < 2e-16 ***
## FirstAuthorFemale1 0.0255 0.0242 1.05 0.29200
## LastAuthorFemale1 0.0367 0.0307 1.20 0.23178
## UniqueAuthors2 -0.1496 0.0772 -1.94 0.05301 .
## UniqueAuthors3 -0.1333 0.0753 -1.77 0.07701 .
## UniqueAuthors4 -0.1181 0.0754 -1.57 0.11729
## UniqueAuthors5 -0.0611 0.0747 -0.82 0.41344
## Year1997 -0.0421 0.0623 -0.68 0.49947
## Year1998 -0.0370 0.0584 -0.63 0.52601
## Year1999 -0.1299 0.0597 -2.17 0.02981 *
```

```

## Year2000          -0.0613      0.0515    -1.19   0.23393
## Year2001          -0.1959      0.0591    -3.31   0.00094 ***
## Year2002          -0.1066      0.0565    -1.89   0.05958 .
## Year2003          -0.1507      0.0551    -2.74   0.00632 **
## Year2004          -0.1894      0.0611    -3.10   0.00197 **
## Year2005          -0.1205      0.0571    -2.11   0.03501 *
## Year2006          -0.1663      0.0551    -3.02   0.00258 **
## Year2007          -0.1816      0.0502    -3.62   0.00031 ***
## Year2008          -0.2195      0.0583    -3.76   0.00018 ***
## Year2009          -0.1056      0.0527    -2.00   0.04542 *
## Year2010          -0.1562      0.0555    -2.81   0.00499 **
## Year2011          -0.1834      0.0549    -3.34   0.00085 ***
## Year2012          -0.1371      0.0573    -2.39   0.01683 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0394, Adjusted R-squared:  0.0232
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1216 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.173  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          7.54e-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 1.067 1          1.033
## LastAuthorFemale  1.036 1          1.018
## Year              1.105 16          1.003

```

Residuals from first and last author



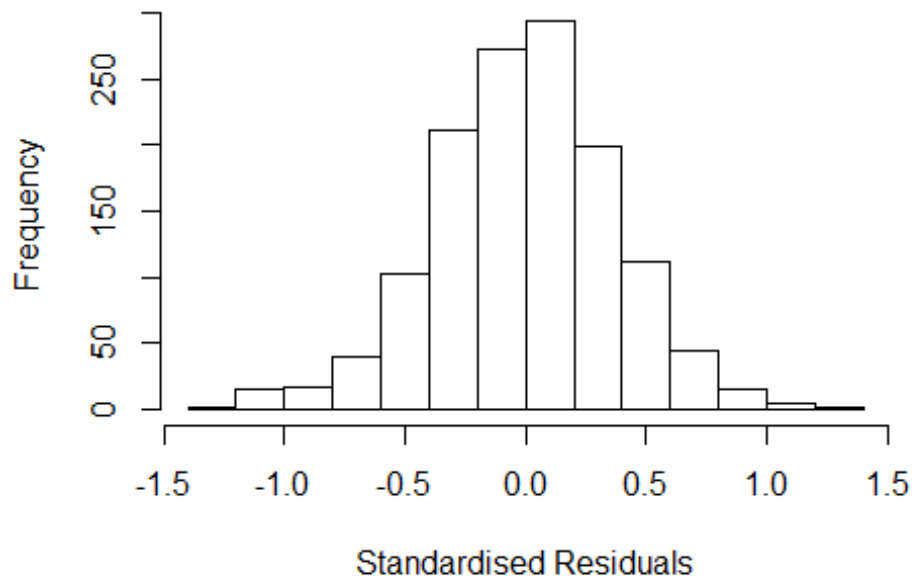
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2307 -0.2404 0.0034 0.2348 1.2209
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2456 0.0386 32.30 < 2e-16 ***
## FirstAuthorFemale1 0.0239 0.0242 0.99 0.32299
## LastAuthorFemale1 0.0342 0.0308 1.11 0.26629
## Year1997 -0.0458 0.0624 -0.73 0.46332
## Year1998 -0.0388 0.0583 -0.67 0.50602
## Year1999 -0.1316 0.0605 -2.18 0.02970 *
## Year2000 -0.0515 0.0518 -0.99 0.32004
## Year2001 -0.1926 0.0602 -3.20 0.00140 **
## Year2002 -0.1118 0.0579 -1.93 0.05363 .
## Year2003 -0.1515 0.0550 -2.75 0.00598 **
## Year2004 -0.1823 0.0615 -2.96 0.00309 **
## Year2005 -0.1085 0.0575 -1.89 0.05940 .
```

```

## Year2006          -0.1464      0.0555   -2.64  0.00843 **
## Year2007          -0.1750      0.0508   -3.44  0.00059 ***
## Year2008          -0.2115      0.0586   -3.61  0.00032 ***
## Year2009          -0.0918      0.0532   -1.73  0.08467 .
## Year2010          -0.1417      0.0552   -2.57  0.01042 *
## Year2011          -0.1719      0.0553   -3.11  0.00193 **
## Year2012          -0.1245      0.0573   -2.17  0.02992 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.0147
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 133 weights are ~= 1. The remaining 1194 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.871  0.952  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      7.54e-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 1.073 1      1.036
## Year              1.073 16      1.002

```

Residuals from first author



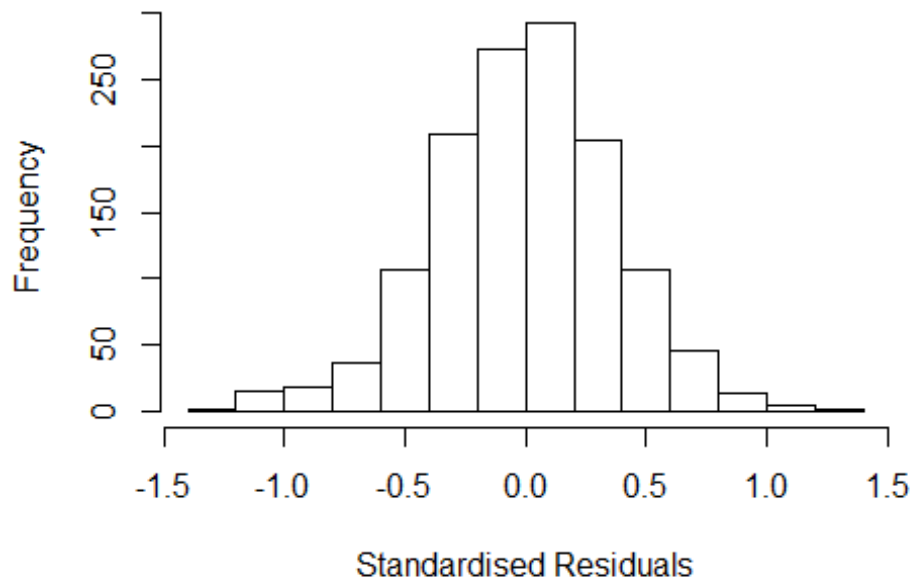
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23702 -0.24008  0.00239  0.23604  1.21833
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2474    0.0388   32.16 < 2e-16 ***
## FirstAuthorFemale1  0.0250    0.0243    1.03  0.30454
## Year1997        -0.0437    0.0626   -0.70  0.48471
## Year1998        -0.0353    0.0582   -0.61  0.54410
## Year1999        -0.1275    0.0605   -2.11  0.03536 *
## Year2000        -0.0488    0.0518   -0.94  0.34631
## Year2001        -0.1910    0.0603   -3.17  0.00158 **
## Year2002        -0.1082    0.0579   -1.87  0.06183 .
## Year2003        -0.1478    0.0550   -2.69  0.00731 **
## Year2004        -0.1805    0.0616   -2.93  0.00344 **
## Year2005        -0.1077    0.0574   -1.88  0.06101 .
## Year2006        -0.1441    0.0555   -2.60  0.00953 **
```

```

## Year2007          -0.1739      0.0511   -3.41  0.00068 ***
## Year2008          -0.2088      0.0586   -3.56  0.00038 ***
## Year2009          -0.0877      0.0530   -1.65  0.09849 .
## Year2010          -0.1388      0.0551   -2.52  0.01193 *
## Year2011          -0.1705      0.0555   -3.07  0.00215 **
## Year2012          -0.1206      0.0573   -2.10  0.03567 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.027, Adjusted R-squared:  0.0144
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1199 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.870  0.952  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      7.54e-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 1.039 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
## -1.21e+00 -2.40e-01 -5.46e-05 2.37e-01 1.22e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2513 0.0379 32.99 < 2e-16 ***
## LastAuthorFemale1 0.0354 0.0310 1.14 0.25283
## Year1997 -0.0459 0.0623 -0.74 0.46169
## Year1998 -0.0392 0.0579 -0.68 0.49819
## Year1999 -0.1283 0.0601 -2.13 0.03298 *
## Year2000 -0.0506 0.0518 -0.98 0.32823
## Year2001 -0.1946 0.0601 -3.23 0.00125 **
## Year2002 -0.1100 0.0577 -1.91 0.05697 .
## Year2003 -0.1516 0.0552 -2.75 0.00611 **
## Year2004 -0.1810 0.0616 -2.94 0.00335 **
## Year2005 -0.1106 0.0576 -1.92 0.05483 .
## Year2006 -0.1465 0.0553 -2.65 0.00813 **
```

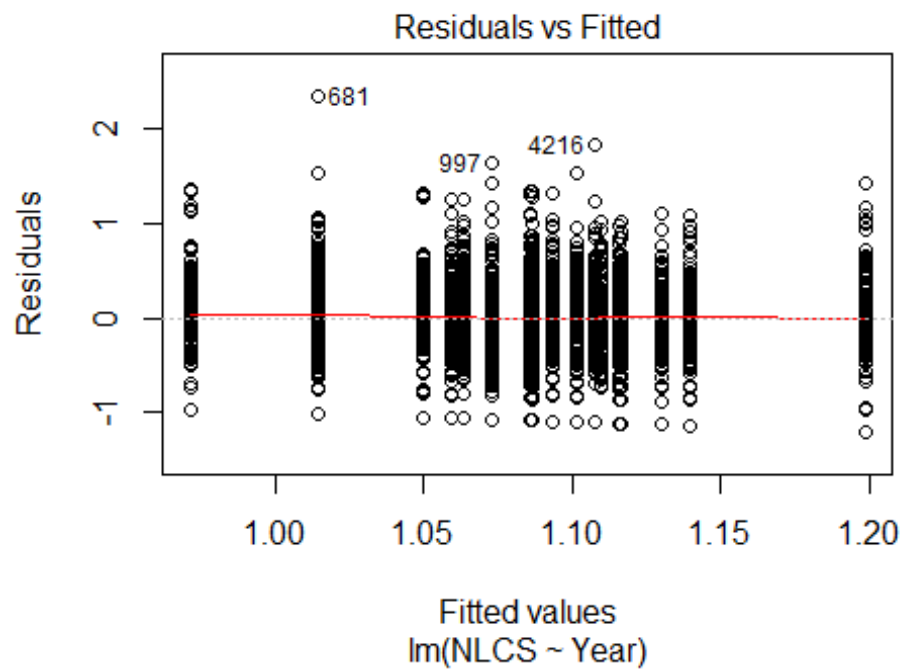
```

## Year2007          -0.1750      0.0506   -3.46  0.00056 ***
## Year2008          -0.2118      0.0587   -3.61  0.00032 ***
## Year2009          -0.0912      0.0532   -1.71  0.08694 .
## Year2010          -0.1433      0.0552   -2.59  0.00960 **
## Year2011          -0.1742      0.0551   -3.16  0.00161 **
## Year2012          -0.1239      0.0574   -2.16  0.03101 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0273, Adjusted R-squared:  0.0146
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 126 weights are ~ = 1. The remaining 1201 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.219  0.871  0.952  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      7.54e-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: 1327"
## [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
## 249 254 237 232 252 241 175 188 210 267 251 241 256 282 317
## 2011 2012
## 358 318
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 170 172 184 151 146 148 126 138 149 201 190 186 186 198 219
## 2011 2012

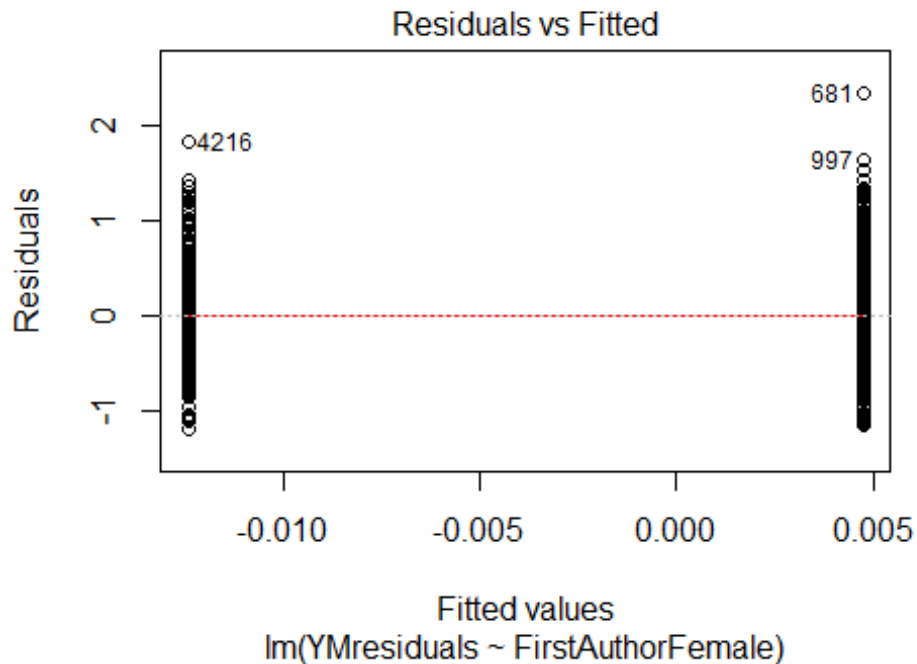
```



```
## 245 214
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 153 157 166 131 128 130 113 125 138 183 176 172 170 181 203
## 2011 2012
## 219 192
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 98, df = 16, p-value = 8e-14
```

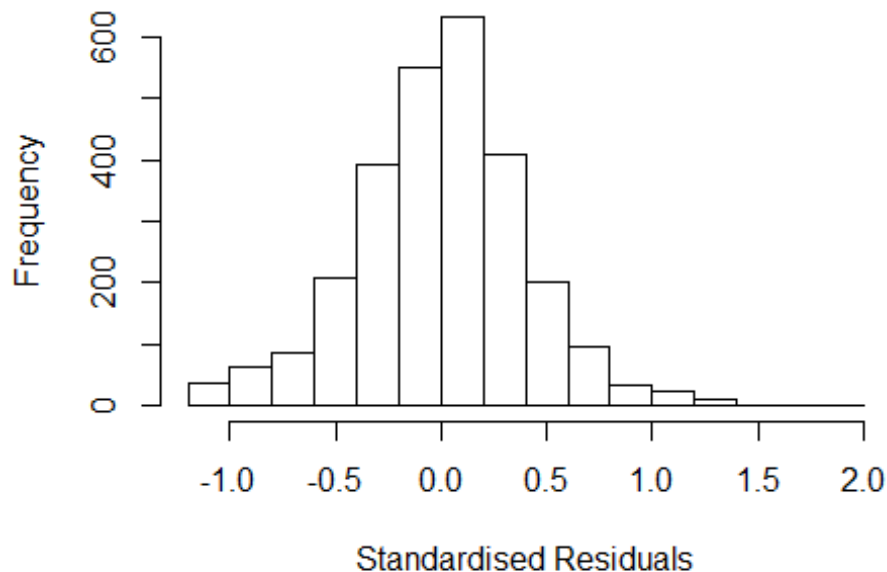


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.6, df = 1, p-value = 0.06
```



```
## [1] "Female first author team size 2018 geometric mean: 5.10926851582627"
## [1] "Male first author team size 2018 geometric mean: 4.51528714224219"
##
## 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: 4.63028669297144"
## [1] "Male last author team size 2018 geometric mean: 4.72538217166001"
##
## 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] "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 1.062 1 1.031
## UniqueAuthors 1.204 4 1.023
## Year 1.279 16 1.008
```

Residuals from first and last author and team size



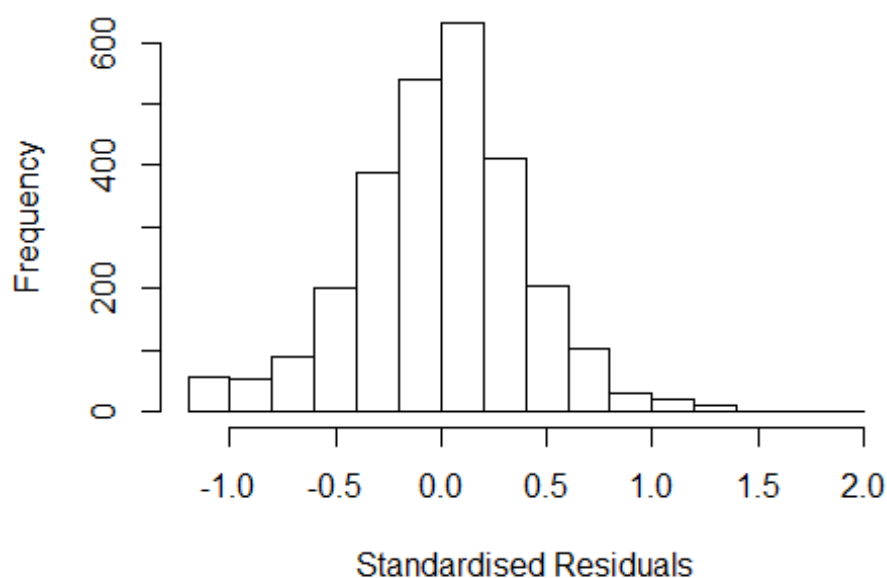
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1933 -0.2424 0.0164 0.2338 1.9292
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02318 0.06912 14.80 < 2e-16 ***
## FirstAuthorFemale1 -0.01834 0.01671 -1.10 0.27239
## LastAuthorFemale1 0.03225 0.02277 1.42 0.15684
## UniqueAuthors2 0.08721 0.06616 1.32 0.18758
## UniqueAuthors3 0.14369 0.06486 2.22 0.02682 *
## UniqueAuthors4 0.16199 0.06475 2.50 0.01241 *
## UniqueAuthors5 0.20632 0.06454 3.20 0.00141 **
## Year1997 -0.04489 0.04622 -0.97 0.33156
## Year1998 -0.13422 0.05616 -2.39 0.01692 *
## Year1999 -0.09758 0.04547 -2.15 0.03197 *
```

```

## Year2000      -0.05723      0.04921      -1.16      0.24500
## Year2001      -0.20313      0.05373      -3.78      0.00016 ***
## Year2002       0.00225      0.04664       0.05      0.96147
## Year2003      -0.10793      0.04773      -2.26      0.02384 *
## Year2004      -0.05816      0.04479      -1.30      0.19422
## Year2005      -0.09961      0.04278      -2.33      0.01998 *
## Year2006      -0.11229      0.04171      -2.69      0.00714 **
## Year2007      -0.05172      0.04036      -1.28      0.20007
## Year2008      -0.04508      0.04149      -1.09      0.27725
## Year2009      -0.05279      0.04086      -1.29      0.19642
## Year2010      -0.09429      0.03954      -2.38      0.01716 *
## Year2011      -0.13096      0.04010      -3.27      0.00110 **
## Year2012      -0.06848      0.04191      -1.63      0.10236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.034, Adjusted R-squared:  0.0262
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 2177 is an outlier with |weight| = 0 ( < 3.7e-05);
## 221 weights are ~= 1. The remaining 2515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.034  0.867  0.952  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           3.65e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "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 1.057 1 1.028
## Year 1.086 16 1.003

```

Residuals from first and last author



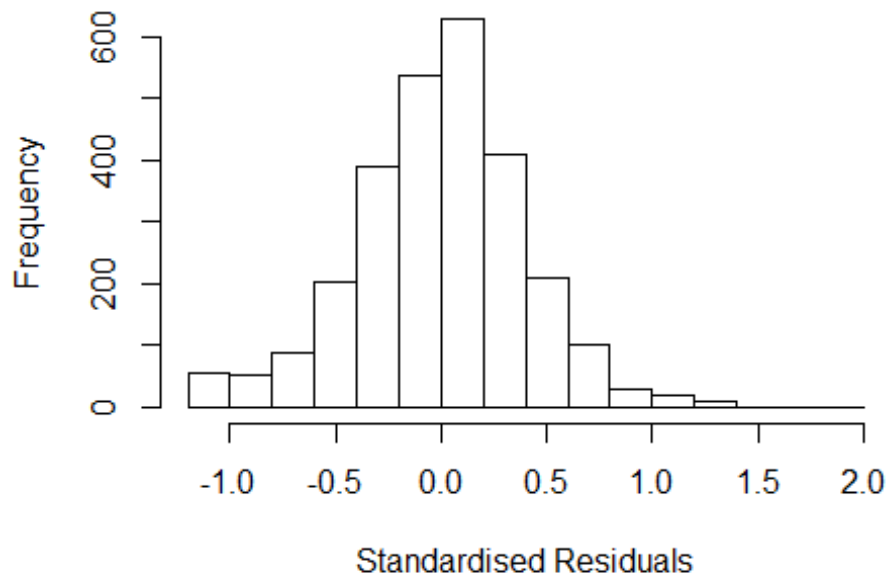
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1554 -0.2441 0.0132 0.2352 1.8558
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1554 0.0314 36.74 < 2e-16 ***
## FirstAuthorFemale1 -0.0154 0.0168 -0.91 0.36031
## LastAuthorFemale1 0.0366 0.0230 1.59 0.11161
## Year1997 -0.0375 0.0458 -0.82 0.41321
## Year1998 -0.1267 0.0564 -2.25 0.02482 *
## Year1999 -0.0880 0.0455 -1.94 0.05297 .
## Year2000 -0.0489 0.0496 -0.99 0.32371
## Year2001 -0.1868 0.0535 -3.49 0.00049 ***
## Year2002 0.0227 0.0464 0.49 0.62446
## Year2003 -0.0978 0.0476 -2.05 0.04010 *
## Year2004 -0.0456 0.0448 -1.02 0.30856
## Year2005 -0.0849 0.0423 -2.01 0.04483 *
```

```

## Year2006          -0.0908      0.0417    -2.17  0.02976 *
## Year2007          -0.0242      0.0399    -0.61  0.54396
## Year2008          -0.0291      0.0412    -0.71  0.48002
## Year2009          -0.0326      0.0401    -0.81  0.41626
## Year2010          -0.0689      0.0389    -1.77  0.07651 .
## Year2011          -0.1026      0.0395    -2.59  0.00953 **
## Year2012          -0.0460      0.0414    -1.11  0.26742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.00911
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2177 is an outlier with |weight| = 0 ( < 3.7e-05);
## 223 weights are ~= 1. The remaining 2513 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0363 0.8660 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          3.65e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



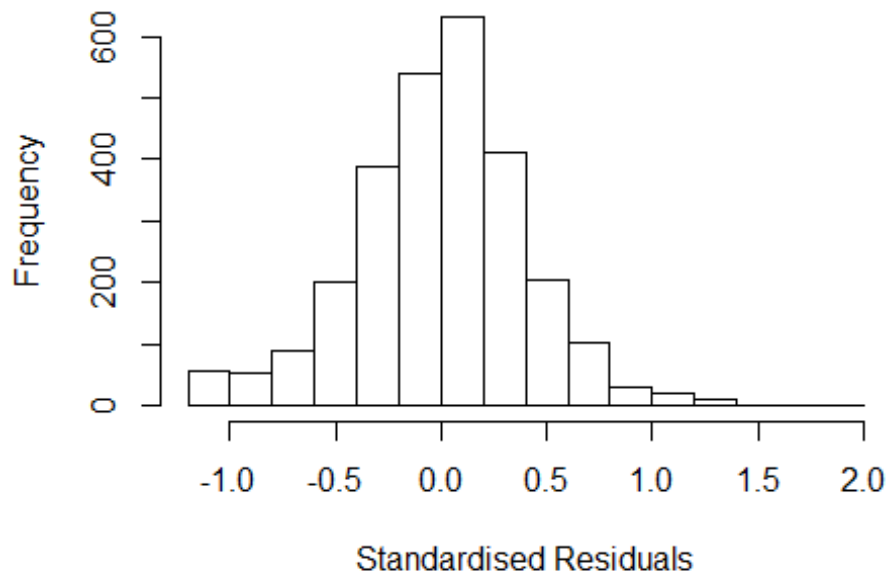
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1575 -0.2425 0.0145 0.2339 1.8505
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1575 0.0315 36.70 < 2e-16 ***
## FirstAuthorFemale1 -0.0151 0.0168 -0.89 0.37123
## Year1997 -0.0370 0.0458 -0.81 0.41929
## Year1998 -0.1253 0.0565 -2.22 0.02655 *
## Year1999 -0.0850 0.0454 -1.87 0.06121 .
## Year2000 -0.0481 0.0496 -0.97 0.33163
## Year2001 -0.1853 0.0537 -3.45 0.00056 ***
## Year2002 0.0260 0.0466 0.56 0.57676
## Year2003 -0.0943 0.0476 -1.98 0.04779 *
## Year2004 -0.0428 0.0449 -0.95 0.34068
## Year2005 -0.0828 0.0424 -1.95 0.05087 .
## Year2006 -0.0884 0.0418 -2.11 0.03470 *
```

```

## Year2007          -0.0224      0.0400   -0.56  0.57506
## Year2008          -0.0281      0.0414   -0.68  0.49744
## Year2009          -0.0280      0.0402   -0.70  0.48503
## Year2010          -0.0660      0.0388   -1.70  0.08941 .
## Year2011          -0.1000      0.0396   -2.53  0.01159 *
## Year2012          -0.0428      0.0415   -1.03  0.30243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.0084
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2177 is an outlier with |weight| = 0 ( < 3.7e-05);
## 219 weights are ~= 1. The remaining 2517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0379 0.8650 0.9530 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      3.65e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



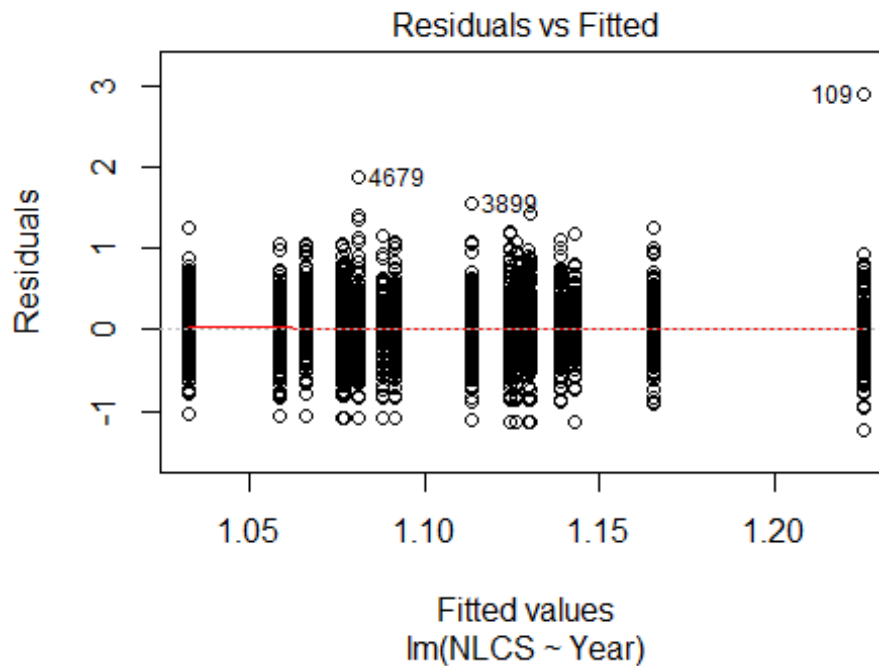
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1515 -0.2482 0.0156 0.2335 1.8442
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1515 0.0310 37.19 <2e-16 ***
## LastAuthorFemale1 0.0364 0.0230 1.58 0.1145
## Year1997 -0.0371 0.0457 -0.81 0.4176
## Year1998 -0.1265 0.0565 -2.24 0.0252 *
## Year1999 -0.0873 0.0454 -1.92 0.0545 .
## Year2000 -0.0489 0.0496 -0.99 0.3246
## Year2001 -0.1865 0.0535 -3.49 0.0005 ***
## Year2002 0.0233 0.0463 0.50 0.6148
## Year2003 -0.0983 0.0477 -2.06 0.0393 *
## Year2004 -0.0460 0.0448 -1.03 0.3044
## Year2005 -0.0852 0.0423 -2.01 0.0441 *
## Year2006 -0.0926 0.0417 -2.22 0.0264 *
```

```

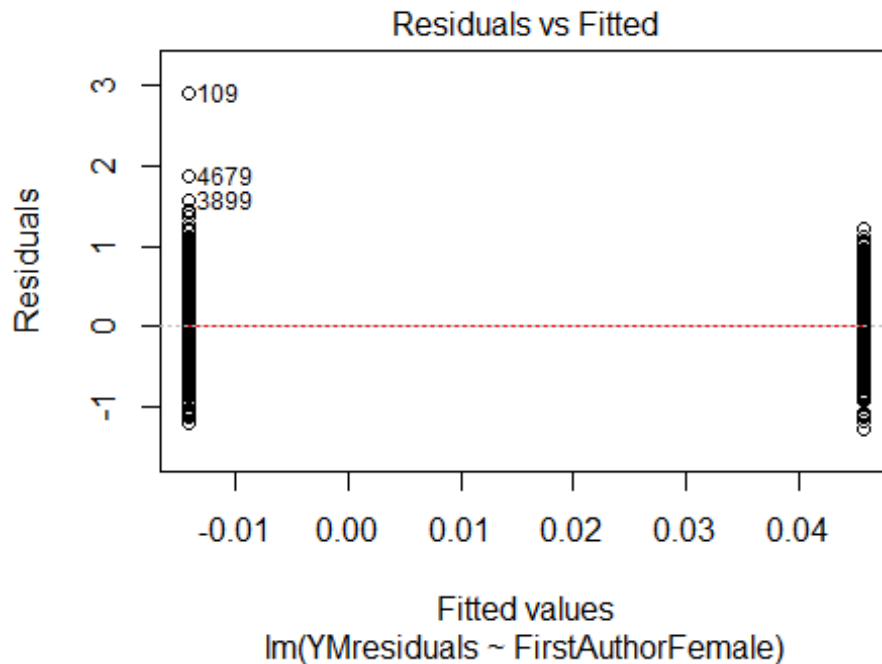
## Year2007          -0.0249      0.0399   -0.62    0.5326
## Year2008          -0.0298      0.0413   -0.72    0.4698
## Year2009          -0.0336      0.0401   -0.84    0.4025
## Year2010          -0.0688      0.0388   -1.77    0.0767 .
## Year2011          -0.1030      0.0395   -2.60    0.0092 **
## Year2012          -0.0464      0.0414   -1.12    0.2627
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.00921
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2177 is an outlier with |weight| = 0 ( < 3.7e-05);
## 218 weights are ~= 1. The remaining 2518 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0353 0.8660 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          3.65e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2737"
## [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
## 317 238 320 270 296 218 214 311 269 302 327 366 323 329 321
## 2011 2012
## 402 354
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 185 148 193 166 161 107 134 204 158 179 190 234 195 219 199

```

```
## 2011 2012
## 256 238
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 159 123 179 150 146 91 121 174 143 157 168 200 162 182 176
## 2011 2012
## 225 211
## [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
```

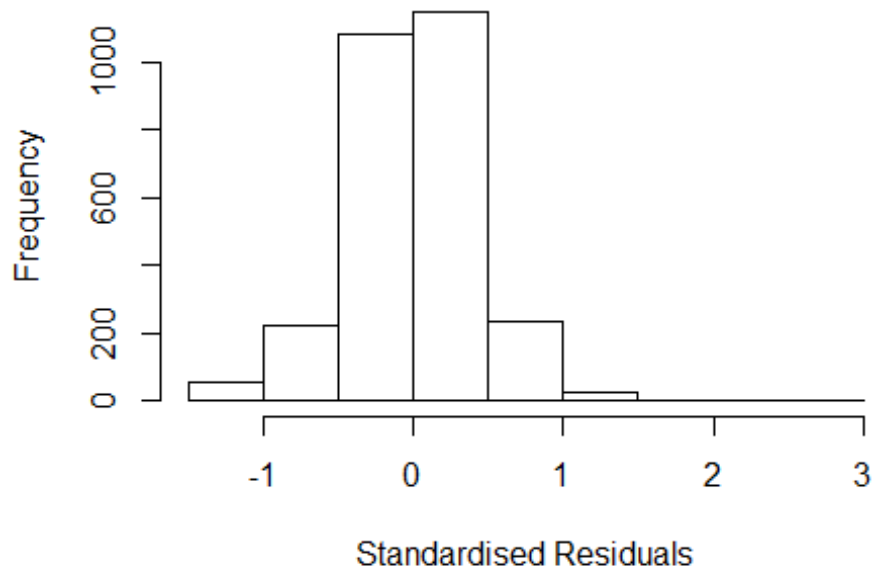


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.8, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.00467826278749"
## [1] "Male first author team size 2018 geometric mean: 3.33244137246508"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.759331977307"
## [1] "Male last author team size 2018 geometric mean: 3.47571223026601"
##
## 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 1.048 1      1.024
## LastAuthorFemale  1.049 1      1.024
## UniqueAuthors    1.160 4      1.019
## Year              1.194 16     1.006
```

Residuals from first and last author and team size



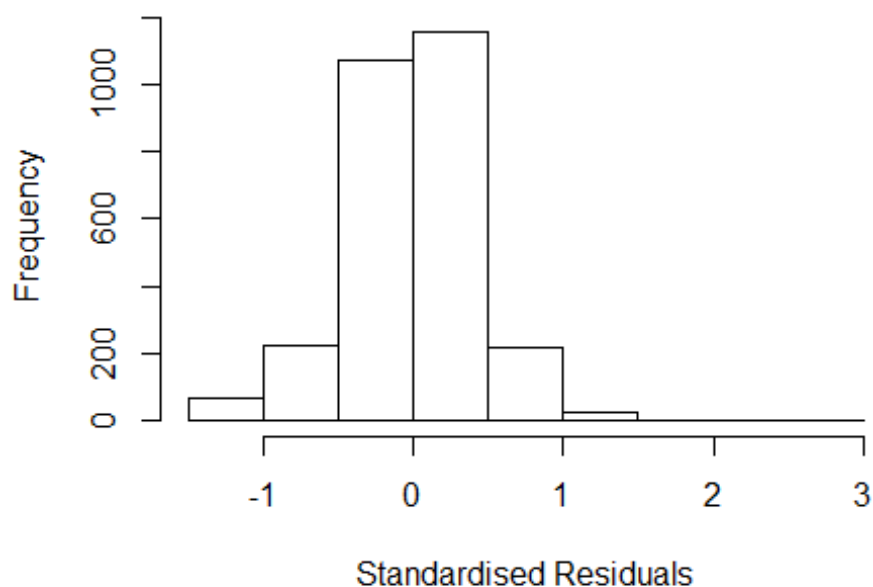
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 109 0011083273 4.117 1996    1606      2      2.92
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29093 -0.24529  0.00611  0.24154  2.91972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1140     0.0464   23.98 < 2e-16 ***
## FirstAuthorFemale1  0.0383     0.0174    2.21  0.02751 *
## LastAuthorFemale1  0.0714     0.0232    3.08  0.00211 **
## UniqueAuthors2     0.0833     0.0372    2.24  0.02511 *
## UniqueAuthors3     0.1171     0.0356    3.29  0.00101 **
## UniqueAuthors4     0.1056     0.0369    2.86  0.00428 **
## UniqueAuthors5     0.1539     0.0368    4.18  3.0e-05 ***
## Year1997          -0.0725     0.0482   -1.51  0.13219
## Year1998          -0.1215     0.0520   -2.33  0.01962 *
## Year1999          -0.0756     0.0450   -1.68  0.09312 .
```

```

## Year2000          -0.0596      0.0426   -1.40   0.16200
## Year2001          -0.0968      0.0547   -1.77   0.07686 .
## Year2002          -0.1375      0.0451   -3.05   0.00231 **
## Year2003          -0.1493      0.0420   -3.55   0.00039 ***
## Year2004          -0.1890      0.0469   -4.03   5.8e-05 ***
## Year2005          -0.1547      0.0464   -3.33   0.00087 ***
## Year2006          -0.1732      0.0438   -3.96   7.8e-05 ***
## Year2007          -0.1501      0.0424   -3.54   0.00040 ***
## Year2008          -0.1571      0.0432   -3.63   0.00028 ***
## Year2009          -0.0949      0.0437   -2.17   0.02994 *
## Year2010          -0.1153      0.0442   -2.61   0.00910 **
## Year2011          -0.1378      0.0419   -3.29   0.00103 **
## Year2012          -0.1482      0.0456   -3.25   0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.0292, Adjusted R-squared:  0.0214
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(50,1949) are outliers with |weight| = 0 ( < 3.6e-05);
## 243 weights are ~= 1. The remaining 2522 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0221 0.8620 0.9510 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          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 1.025 1          1.012
## LastAuthorFemale 1.036 1          1.018
## Year             1.059 16          1.002

```

Residuals from first and last author



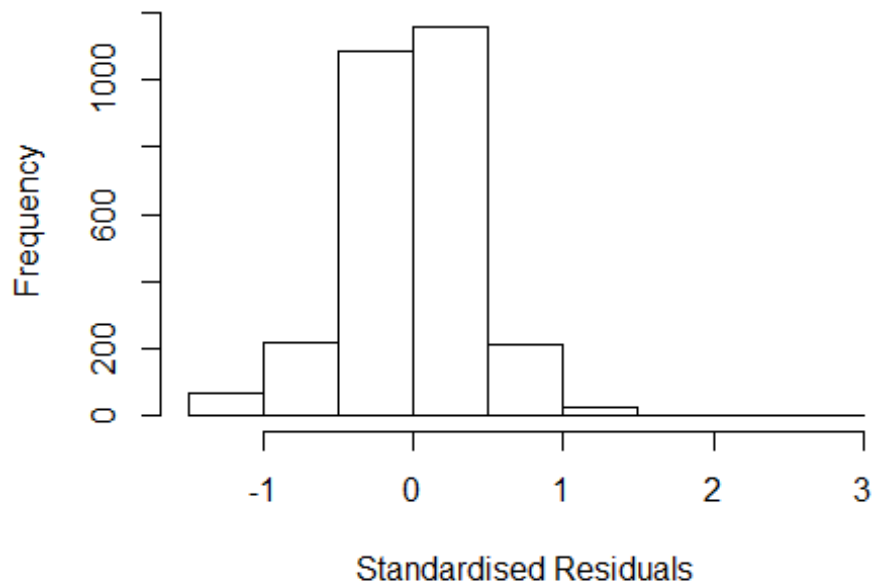
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 109 0011083273 4.117 1996      1606      2      2.903
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33746 -0.24285  0.00608  0.24377  2.90262
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2144     0.0331   36.67 < 2e-16 ***
## FirstAuthorFemale1  0.0476     0.0172    2.77  0.00572 **
## LastAuthorFemale1  0.0755     0.0232    3.26  0.00112 **
## Year1997          -0.0772     0.0480   -1.61  0.10784
## Year1998          -0.1279     0.0518   -2.47  0.01365 *
## Year1999          -0.0853     0.0445   -1.91  0.05575 .
## Year2000          -0.0603     0.0425   -1.42  0.15595
## Year2001          -0.0980     0.0548   -1.79  0.07400 .
## Year2002          -0.1380     0.0450   -3.07  0.00217 **
## Year2003          -0.1480     0.0423   -3.50  0.00047 ***
## Year2004          -0.1858     0.0467   -3.98  7.2e-05 ***
## Year2005          -0.1476     0.0463   -3.19  0.00146 **
```

```

## Year2006          -0.1695      0.0439   -3.86  0.00012 ***
## Year2007          -0.1421      0.0424   -3.35  0.00081 ***
## Year2008          -0.1536      0.0435   -3.53  0.00042 ***
## Year2009          -0.0867      0.0435   -1.99  0.04642 *
## Year2010          -0.1128      0.0439   -2.57  0.01021 *
## Year2011          -0.1284      0.0420   -3.05  0.00228 **
## Year2012          -0.1460      0.0455   -3.21  0.00136 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.019, Adjusted R-squared:  0.0126
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(50,1949) are outliers with |weight| = 0 ( < 3.6e-05);
## 251 weights are ~ = 1. The remaining 2514 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0207 0.8620 0.9510 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          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



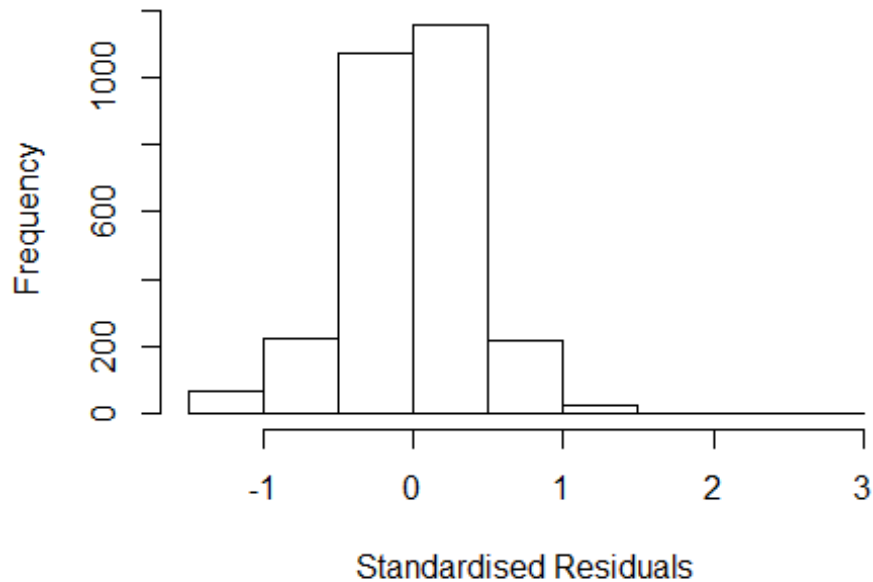
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 109 0011083273 4.117 1996      1606      2      2.903
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26572 -0.24531  0.00707  0.24254  2.90254
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2145     0.0332   36.62 < 2e-16 ***
## FirstAuthorFemale1  0.0513     0.0172    2.98  0.00295 **
## Year1997        -0.0732     0.0480   -1.52  0.12741
## Year1998        -0.1177     0.0517   -2.28  0.02284 *
## Year1999        -0.0798     0.0444   -1.79  0.07280 .
## Year2000        -0.0541     0.0424   -1.28  0.20208
## Year2001        -0.0904     0.0548   -1.65  0.09929 .
## Year2002        -0.1290     0.0447   -2.88  0.00395 **
## Year2003        -0.1403     0.0424   -3.31  0.00094 ***
## Year2004        -0.1751     0.0470   -3.73  0.00020 ***
## Year2005        -0.1376     0.0460   -2.99  0.00278 **
## Year2006        -0.1610     0.0439   -3.67  0.00025 ***
```

```

## Year2007          -0.1323      0.0424   -3.12  0.00184 **
## Year2008          -0.1441      0.0436   -3.31  0.00095 ***
## Year2009          -0.0770      0.0435   -1.77  0.07685 .
## Year2010          -0.1015      0.0438   -2.32  0.02060 *
## Year2011          -0.1210      0.0420   -2.88  0.00401 **
## Year2012          -0.1359      0.0455   -2.99  0.00285 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.00871
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(50,1949) are outliers with |weight| = 0 ( < 3.6e-05);
## 261 weights are ~= 1. The remaining 2504 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0242 0.8620 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      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 1.036 1          1.018
## Year          1.036 16          1.001

```

Residuals from last author



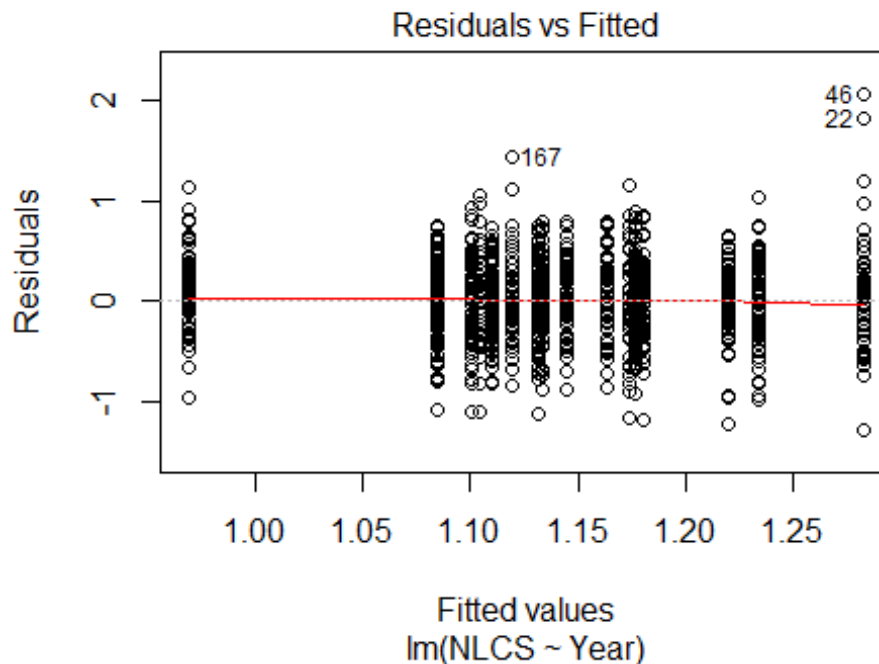
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 109 0011083273 4.117 1996      1606      2      2.903
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.30505 -0.24795  0.00705  0.24250  2.89148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2255     0.0329   37.23 < 2e-16 ***
## LastAuthorFemale1  0.0795     0.0232    3.43 0.00061 ***
## Year1997        -0.0797     0.0479   -1.66 0.09636 .
## Year1998        -0.1286     0.0517   -2.49 0.01288 *
## Year1999        -0.0848     0.0446   -1.90 0.05741 .
## Year2000        -0.0603     0.0428   -1.41 0.15938
## Year2001        -0.0982     0.0549   -1.79 0.07388 .
## Year2002        -0.1400     0.0451   -3.10 0.00194 **
## Year2003        -0.1491     0.0425   -3.51 0.00046 ***
## Year2004        -0.1873     0.0469   -4.00 6.6e-05 ***
## Year2005        -0.1488     0.0464   -3.21 0.00135 **
## Year2006        -0.1714     0.0440   -3.90 9.9e-05 ***
```

```

## Year2007          -0.1402      0.0425   -3.30  0.00099 ***
## Year2008          -0.1523      0.0436   -3.49  0.00049 ***
## Year2009          -0.0859      0.0438   -1.96  0.04976 *
## Year2010          -0.1146      0.0440   -2.60  0.00924 **
## Year2011          -0.1263      0.0422   -2.99  0.00279 **
## Year2012          -0.1446      0.0457   -3.16  0.00158 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.0102
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(50,1949) are outliers with |weight| = 0 ( < 3.6e-05);
## 249 weights are ~ = 1. The remaining 2516 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0248 0.8620 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      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: 2767"
## [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
##   86   99   89  103  104   87  104   90  100  112  117  110  141  150  170
## 2011 2012
##  168  133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   54   48   46   51   63   42   68   58   70   85   81   75   97  122  110

```

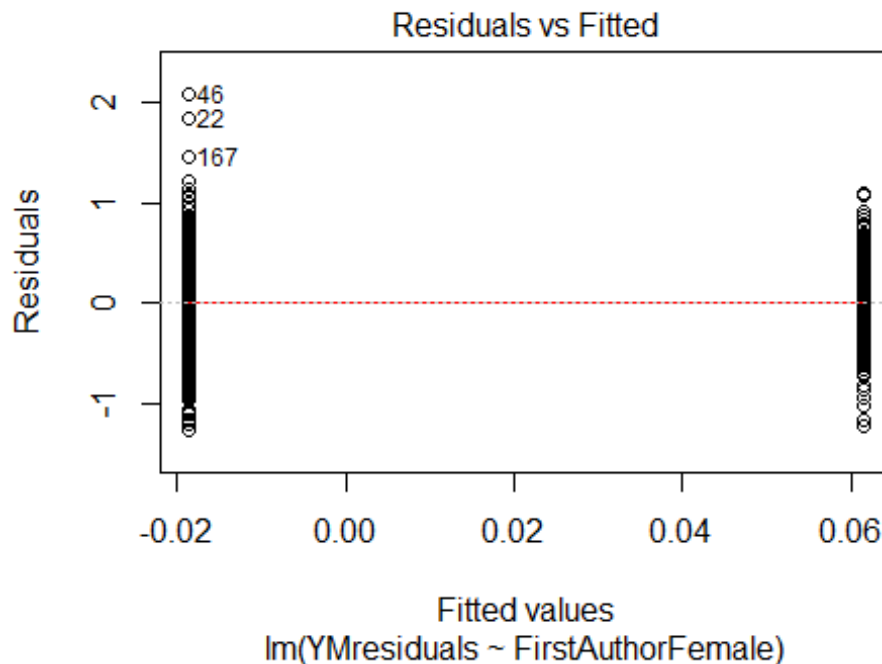
```
## 2011 2012
## 128 95
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 42 38 46 54 36 58 45 61 74 69 63 84 104 97
## 2011 2012
## 103 68
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## 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 = 1.6, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 4.36640274871067"
## [1] "Male first author team size 2018 geometric mean: 4.82074989068677"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

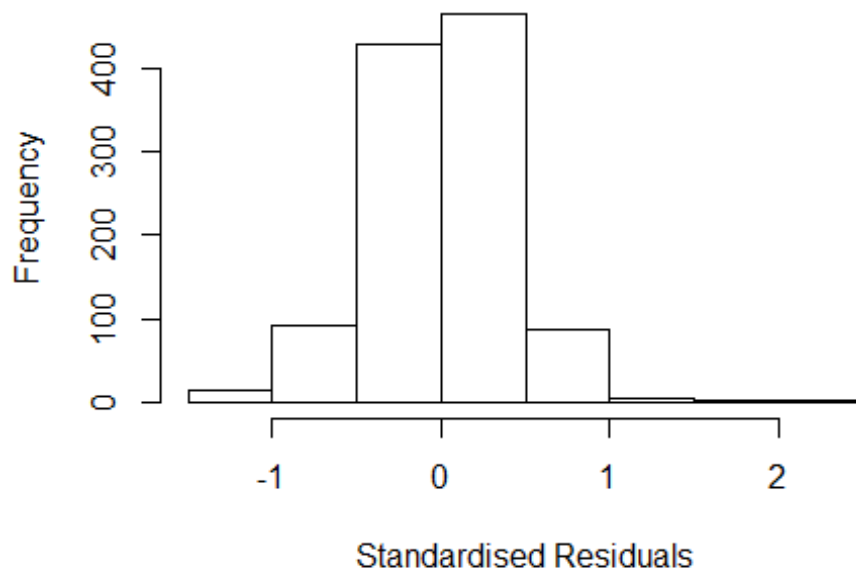
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 300, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.63780974798437"
## [1] "Male last author team size 2018 geometric mean: 4.61810636942053"

## 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 = 270, 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      1.041
## LastAuthorFemale  1.137 1      1.066
## UniqueAuthors    1.581 4      1.059
## Year              1.725 16     1.017
```

Residuals from first and last author and team size



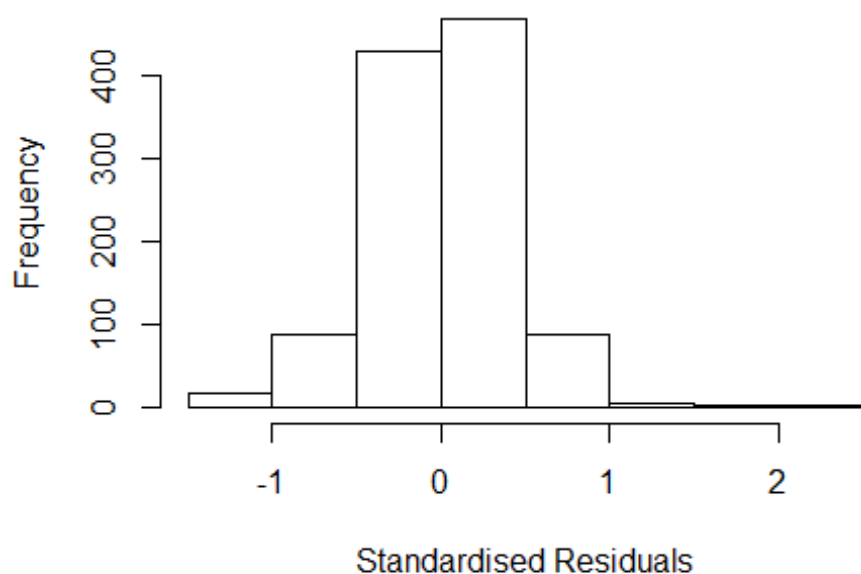
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2017 -0.2440 0.0167 0.2356 2.2149
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12905 0.09120 12.38 <2e-16 ***
## FirstAuthorFemale1 0.05254 0.02718 1.93 0.053 .
## LastAuthorFemale1 0.03668 0.03172 1.16 0.248
## UniqueAuthors2 0.01900 0.06528 0.29 0.771
## UniqueAuthors3 0.04419 0.06306 0.70 0.484
## UniqueAuthors4 0.05401 0.06517 0.83 0.407
## UniqueAuthors5 0.09832 0.06397 1.54 0.125
## Year1997 -0.10610 0.10476 -1.01 0.311
## Year1998 -0.01209 0.11181 -0.11 0.914
## Year1999 0.00415 0.09725 0.04 0.966
```

```

## Year2000      -0.03596    0.08904   -0.40    0.686
## Year2001      0.00111    0.11247    0.01    0.992
## Year2002     -0.01994    0.08879   -0.22    0.822
## Year2003      0.05447    0.09342    0.58    0.560
## Year2004     -0.01407    0.08553   -0.16    0.869
## Year2005     -0.08324    0.08118   -1.03    0.305
## Year2006     -0.07488    0.08955   -0.84    0.403
## Year2007     -0.05530    0.08599   -0.64    0.520
## Year2008      0.00329    0.08538    0.04    0.969
## Year2009     -0.07690    0.08072   -0.95    0.341
## Year2010     -0.09079    0.08123   -1.12    0.264
## Year2011     -0.09607    0.08278   -1.16    0.246
## Year2012     -0.15695    0.08810   -1.78    0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0249, Adjusted R-squared:  0.00485
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(12,26) are outliers with |weight| = 0 ( < 9.1e-05);
## 84 weights are ~= 1. The remaining 1008 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.164  0.873  0.953  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           9.14e-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 1.057 1 1.028
## LastAuthorFemale 1.113 1 1.055
## Year 1.158 16 1.005

```


Residuals from first and last author



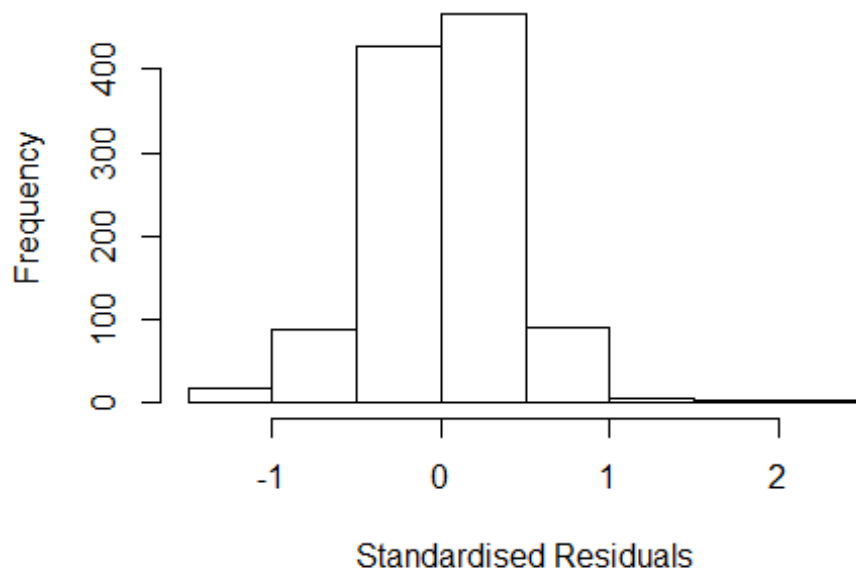
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.235 -0.240 0.015 0.237 2.171
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17298 0.07259 16.16 <2e-16 ***
## FirstAuthorFemale1 0.05782 0.02673 2.16 0.031 *
## LastAuthorFemale1 0.03299 0.03190 1.03 0.301
## Year1997 -0.10638 0.10318 -1.03 0.303
## Year1998 -0.02291 0.11009 -0.21 0.835
## Year1999 -0.00207 0.09676 -0.02 0.983
## Year2000 -0.04175 0.08842 -0.47 0.637
## Year2001 0.00246 0.11099 0.02 0.982
## Year2002 -0.01567 0.08742 -0.18 0.858
## Year2003 0.06179 0.09092 0.68 0.497
## Year2004 -0.00707 0.08372 -0.08 0.933
## Year2005 -0.07242 0.08026 -0.90 0.367
```

```

## Year2006      -0.07514    0.08884   -0.85    0.398
## Year2007      -0.04557    0.08532   -0.53    0.593
## Year2008       0.01370    0.08442    0.16    0.871
## Year2009      -0.07226    0.08036   -0.90    0.369
## Year2010      -0.07201    0.08029   -0.90    0.370
## Year2011      -0.08901    0.08241   -1.08    0.280
## Year2012      -0.13722    0.08797   -1.56    0.119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.0194, Adjusted R-squared:  0.00303
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(12,26) are outliers with |weight| = 0 ( < 9.1e-05);
## 94 weights are ~= 1. The remaining 998 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.195  0.873   0.953   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      9.14e-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 1.047 1          1.023
## Year              1.047 16          1.001

```

Residuals from first author



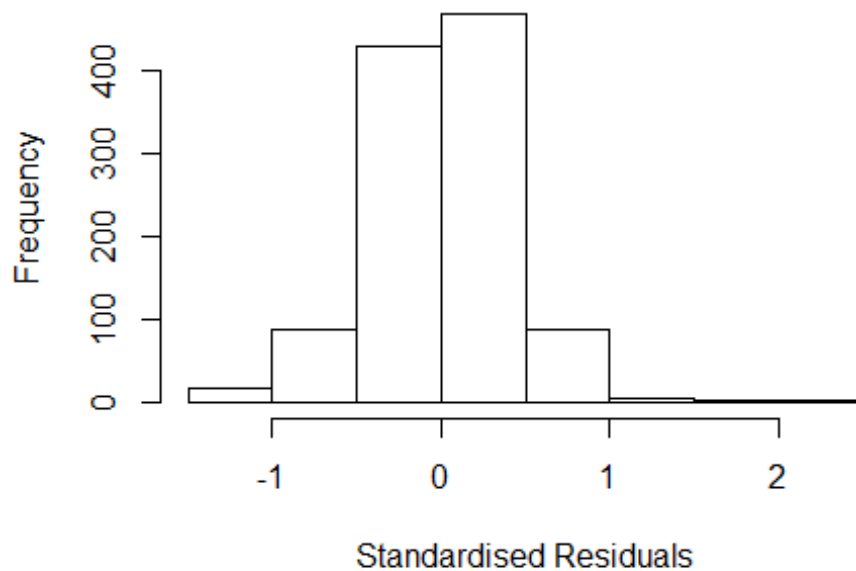
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2390 -0.2431 0.0124 0.2378 2.1662
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17782 0.07221 16.31 <2e-16 ***
## FirstAuthorFemale1 0.05708 0.02678 2.13 0.033 *
## Year1997 -0.10827 0.10300 -1.05 0.293
## Year1998 -0.02014 0.11003 -0.18 0.855
## Year1999 -0.00506 0.09686 -0.05 0.958
## Year2000 -0.04123 0.08817 -0.47 0.640
## Year2001 -0.00162 0.11067 -0.01 0.988
## Year2002 -0.01751 0.08716 -0.20 0.841
## Year2003 0.06119 0.09083 0.67 0.501
## Year2004 -0.00848 0.08358 -0.10 0.919
## Year2005 -0.07035 0.08013 -0.88 0.380
## Year2006 -0.07594 0.08881 -0.86 0.393
```

```

## Year2007          -0.04649    0.08525   -0.55    0.586
## Year2008          0.01582    0.08393    0.19    0.851
## Year2009         -0.06862    0.08000   -0.86    0.391
## Year2010         -0.06793    0.07985   -0.85    0.395
## Year2011         -0.08415    0.08179   -1.03    0.304
## Year2012         -0.13645    0.08771   -1.56    0.120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0183, Adjusted R-squared:  0.00276
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(12,26) are outliers with |weight| = 0 ( < 9.1e-05);
## 102 weights are ~= 1. The remaining 990 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200  0.873   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      9.14e-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 1.104 1         1.051
## Year             1.104 16         1.003

```

Residuals from last author



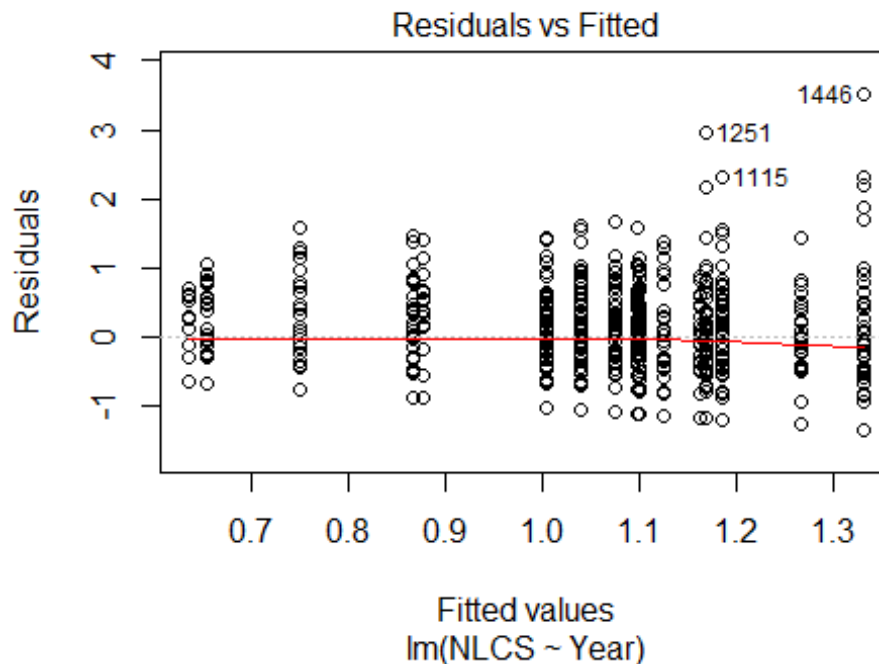
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2493 -0.2426 0.0116 0.2430 2.1623
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18172 0.07227 16.35 <2e-16 ***
## LastAuthorFemale1 0.03115 0.03214 0.97 0.33
## Year1997 -0.10725 0.10335 -1.04 0.30
## Year1998 -0.00854 0.10985 -0.08 0.94
## Year1999 -0.00389 0.09646 -0.04 0.97
## Year2000 -0.03584 0.08874 -0.40 0.69
## Year2001 0.00374 0.11175 0.03 0.97
## Year2002 -0.01084 0.08705 -0.12 0.90
## Year2003 0.06756 0.09124 0.74 0.46
## Year2004 -0.00715 0.08352 -0.09 0.93
## Year2005 -0.06533 0.08019 -0.81 0.42
## Year2006 -0.06982 0.08916 -0.78 0.43
```

```

## Year2007      -0.03712    0.08547   -0.43    0.66
## Year2008      0.02433    0.08414    0.29    0.77
## Year2009     -0.05963    0.08008   -0.74    0.46
## Year2010     -0.06487    0.08041   -0.81    0.42
## Year2011     -0.08158    0.08209   -0.99    0.32
## Year2012     -0.13284    0.08780   -1.51    0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.0153, Adjusted R-squared:  -0.000285
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(12,26) are outliers with |weight| = 0 ( < 9.1e-05);
## 106 weights are ~= 1. The remaining 986 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200  0.873   0.951   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      9.14e-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: 1094"
## [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
##   73   79   55   52   55   58   43   33   23   34   58   94   80  106  105
## 2011 2012
##  103   93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   39   34   30   28   41   25   23   15   24   39   66   53   66   59

```

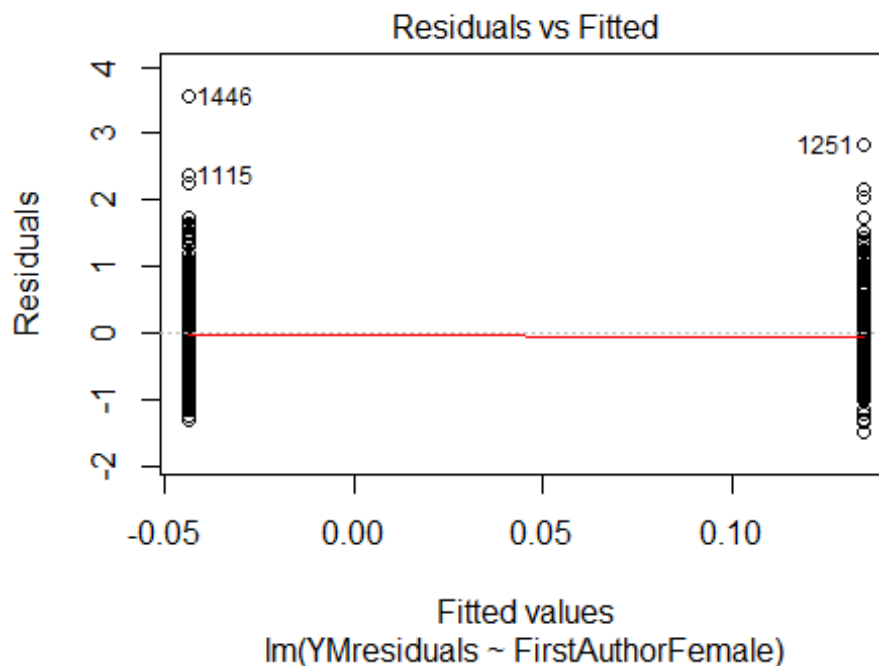
```
## 2011 2012
## 57 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 35 30 24 20 33 20 22 13 21 33 55 40 58 41
## 2011 2012
## 48 45
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 3e-04
```



```
##
## 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: 2.45064105403991"
## [1] "Male first author team size 2018 geometric mean: 2.18017853769217"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

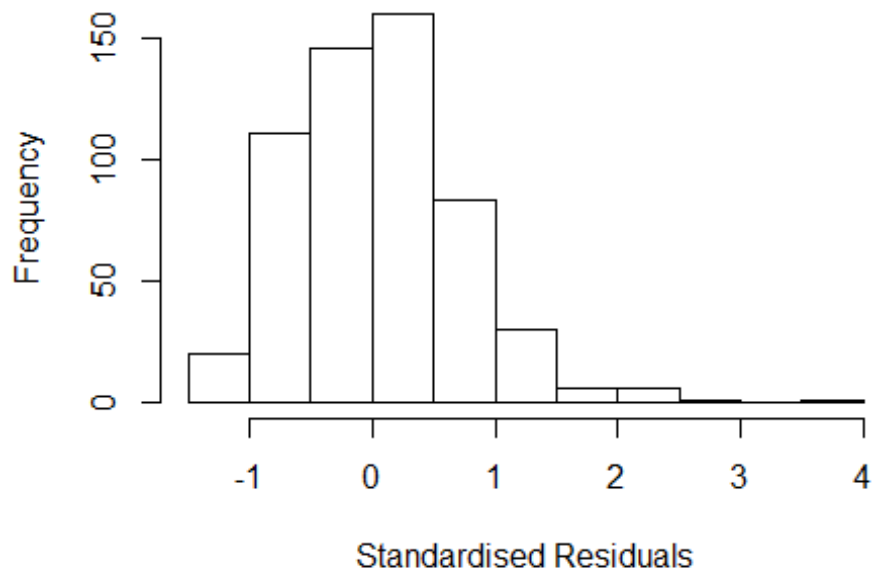
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 350, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.16558669720591"
## [1] "Male last author team size 2018 geometric mean: 2.29337508921497"

## 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 = 0.8
## 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 1.408 1      1.187
## LastAuthorFemale  1.379 1      1.174
## UniqueAuthors    1.515 4      1.053
## Year              1.742 16     1.017
```


Residuals from first and last author and team size



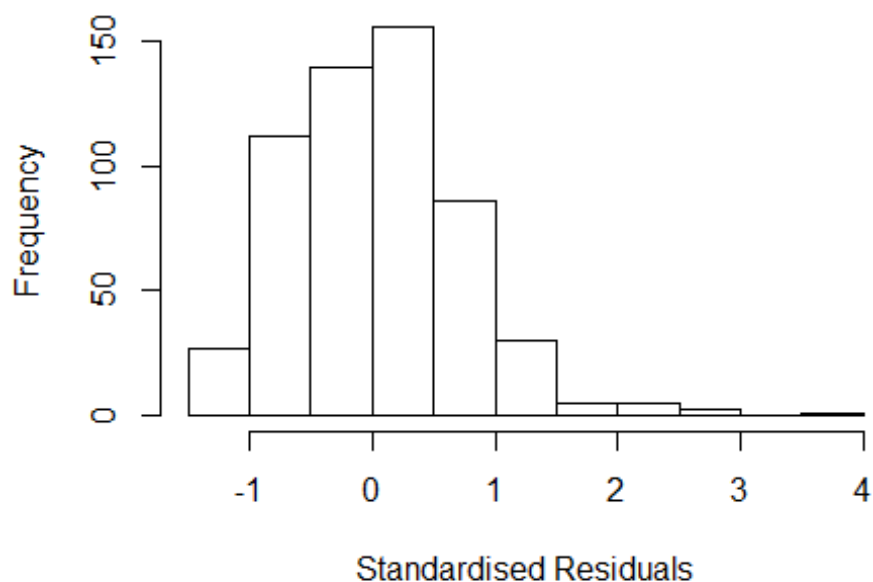
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1251 80052070344 4.129 2011      1700      1      2.713
## 1446 84861815485 4.843 2012      1700      1      3.728
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2397 -0.4650  0.0144  0.4474  3.7277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9521    0.1454    6.55 1.4e-10 ***
## FirstAuthorFemale1 0.2255    0.0766    2.94 0.00338 **
## LastAuthorFemale1 -0.0451    0.0806   -0.56 0.57619
## UniqueAuthors2    0.1182    0.0702    1.68 0.09279 .
## UniqueAuthors3    0.2748    0.0782    3.51 0.00048 ***
## UniqueAuthors4    0.3389    0.1329    2.55 0.01105 *
## UniqueAuthors5    0.2425    0.1811    1.34 0.18114
## Year1997        -0.2542    0.1844   -1.38 0.16862
## Year1998        -0.3302    0.1810   -1.82 0.06873 .
```

```

## Year1999          0.0197      0.2165      0.09  0.92757
## Year2000          0.0576      0.1910      0.30  0.76293
## Year2001         -0.4040      0.1826     -2.21  0.02734 *
## Year2002         -0.0367      0.1870     -0.20  0.84432
## Year2003         -0.2451      0.2024     -1.21  0.22630
## Year2004         -0.4890      0.2124     -2.30  0.02170 *
## Year2005          0.0186      0.1810      0.10  0.91803
## Year2006          0.0136      0.1698      0.08  0.93629
## Year2007         -0.1361      0.1686     -0.81  0.41982
## Year2008         -0.0417      0.1711     -0.24  0.80733
## Year2009         -0.0513      0.1623     -0.32  0.75215
## Year2010         -0.0383      0.1867     -0.21  0.83752
## Year2011         -0.1005      0.1809     -0.56  0.57864
## Year2012         -0.1116      0.1952     -0.57  0.56775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.678
## Multiple R-squared:  0.0943, Adjusted R-squared:  0.0574
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 0.00018);
## 53 weights are ~= 1. The remaining 510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0727 0.8710 0.9470 0.9100 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      1.77e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.346 1 1.160
## LastAuthorFemale 1.327 1 1.152
## Year 1.194 16 1.006

```

Residuals from first and last author



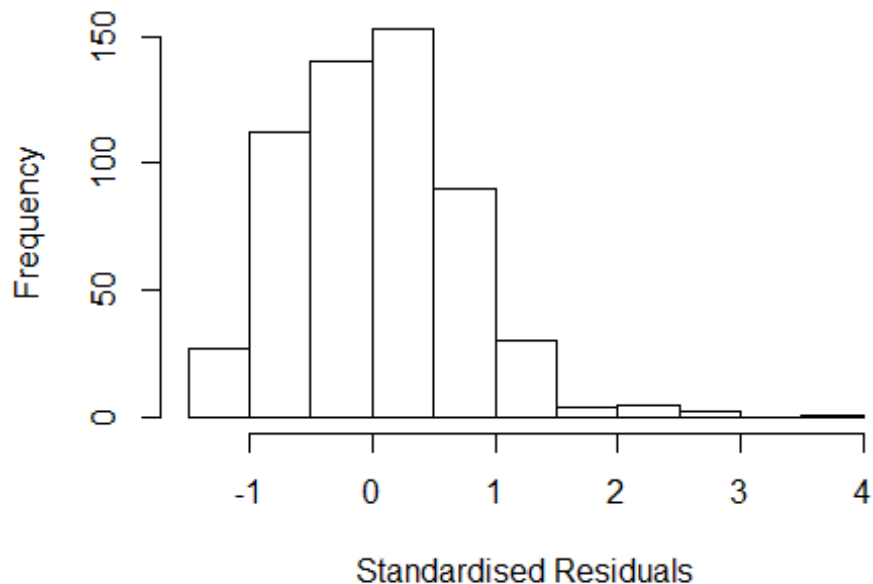
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1251 80052070344 4.129 2011      1700      1      2.867
## 1389 84866505340 3.530 2012      1700      2      2.556
## 1446 84861815485 4.843 2012      1700      1      3.869
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + 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 -0.4735  0.0183  0.4616  3.8692
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.01995    0.14095   7.24 1.6e-12 ***
## FirstAuthorFemale1  0.24970    0.07671   3.25  0.0012 **
## LastAuthorFemale1 -0.05813    0.08032  -0.72  0.4696
## Year1997          -0.25993    0.18175  -1.43  0.1532
## Year1998          -0.36370    0.18238  -1.99  0.0466 *
## Year1999           0.00823    0.21174   0.04  0.9690
## Year2000           0.14661    0.18580   0.79  0.4304
## Year2001          -0.36623    0.18207  -2.01  0.0448 *
## Year2002          -0.00336    0.18430  -0.02  0.9854
## Year2003          -0.23681    0.21327  -1.11  0.2673
```

```

## Year2004          -0.42405      0.20223      -2.10      0.0365 *
## Year2005           0.07069      0.18032       0.39      0.6952
## Year2006           0.03249      0.16908       0.19      0.8477
## Year2007          -0.07728      0.16649      -0.46      0.6427
## Year2008           0.01660      0.17166       0.10      0.9230
## Year2009          -0.02100      0.16153      -0.13      0.8966
## Year2010           0.02930      0.18372       0.16      0.8734
## Year2011          -0.00742      0.18016      -0.04      0.9672
## Year2012          -0.04615      0.19913      -0.23      0.8168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.069, Adjusted R-squared:  0.0383
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 0.00018);
## 46 weights are ~= 1. The remaining 517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0425  0.8700  0.9480  0.9110  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          1.77e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1          1000          200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## 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 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1251 80052070344 4.129 2011      1700      1      2.867
## 1389 84866505340 3.530 2012      1700      2      2.556
## 1446 84861815485 4.843 2012      1700      1      3.869
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2732 -0.4690  0.0163  0.4663  3.8713
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01859    0.14054   7.25 1.5e-12 ***
## FirstAuthorFemale1 0.22313    0.06891   3.24 0.0013 **
## Year1997      -0.27147    0.18066  -1.50 0.1335
## Year1998      -0.36697    0.18175  -2.02 0.0440 *
## Year1999       0.00789    0.21159   0.04 0.9703
## Year2000       0.13872    0.18524   0.75 0.4543
## Year2001      -0.37140    0.18119  -2.05 0.0409 *
## Year2002      -0.00339    0.18363  -0.02 0.9853
## Year2003      -0.23702    0.21285  -1.11 0.2660
## Year2004      -0.43511    0.20311  -2.14 0.0326 *
```

```

## Year2005          0.07566    0.17889    0.42    0.6725
## Year2006          0.02376    0.16903    0.14    0.8883
## Year2007         -0.07959    0.16650   -0.48    0.6329
## Year2008          0.01043    0.17161    0.06    0.9516
## Year2009         -0.02509    0.16091   -0.16    0.8761
## Year2010          0.03151    0.18332    0.17    0.8636
## Year2011         -0.01542    0.17974   -0.09    0.9317
## Year2012         -0.04685    0.19936   -0.24    0.8143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.689
## Multiple R-squared:  0.0683, Adjusted R-squared:  0.0393
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 0.00018);
## 48 weights are ~= 1. The remaining 515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0362 0.8690 0.9450 0.9110 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          1.77e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1          1000          200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## 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
##
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId NLCS Year OneField Fields residuals
## 1251 80052070344 4.129 2011      1700      1      2.867
## 1389 84866505340 3.530 2012      1700      2      2.556
## 1446 84861815485 4.843 2012      1700      1      3.869
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,

```

```

##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.1969 -0.4618  0.0153  0.4557  3.8136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0921     0.1421   7.69   7e-14 ***
## LastAuthorFemale1  0.0726     0.0734   0.99   0.323
## Year1997         -0.3063     0.1838  -1.67   0.096 .
## Year1998         -0.3901     0.1858  -2.10   0.036 *
## Year1999         -0.0368     0.2121  -0.17   0.862
## Year2000          0.0826     0.1885   0.44   0.662
## Year2001         -0.4316     0.1835  -2.35   0.019 *
## Year2002         -0.0237     0.1844  -0.13   0.898
## Year2003         -0.2797     0.2223  -1.26   0.209
## Year2004         -0.4961     0.2028  -2.45   0.015 *
## Year2005          0.0532     0.1799   0.30   0.767
## Year2006         -0.0358     0.1730  -0.21   0.836
## Year2007         -0.1126     0.1714  -0.66   0.512
## Year2008         -0.0218     0.1772  -0.12   0.902
## Year2009         -0.0695     0.1626  -0.43   0.669
## Year2010          0.0321     0.1868   0.17   0.863
## Year2011         -0.0403     0.1780  -0.23   0.821
## Year2012         -0.0628     0.2020  -0.31   0.756
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.692
## Multiple R-squared:  0.0515, Adjusted R-squared:  0.022
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 558 is an outlier with |weight| = 0 ( < 0.00018);
## 47 weights are ~= 1. The remaining 516 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0097 0.8780 0.9490 0.9100 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          1.77e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev mts compute.rd
##      0          1000          0
##      psi          subsampling          cov

```

```

##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 564"
## [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]"
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    1    3    2    5    1    7    10   14    6   11    9   16   10
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    1    2    0    2    1    2    4   10    2    8    8   10    3
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    1    2    0    2    1    2    4    8    2    7    7    9    3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.87298334620742"
## [1] "Male first author team size 2018 geometric mean: 2.61099908058295"

## 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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.46410161513775"
## [1] "Male last author team size 2018 geometric mean: 2.68485274128848"

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

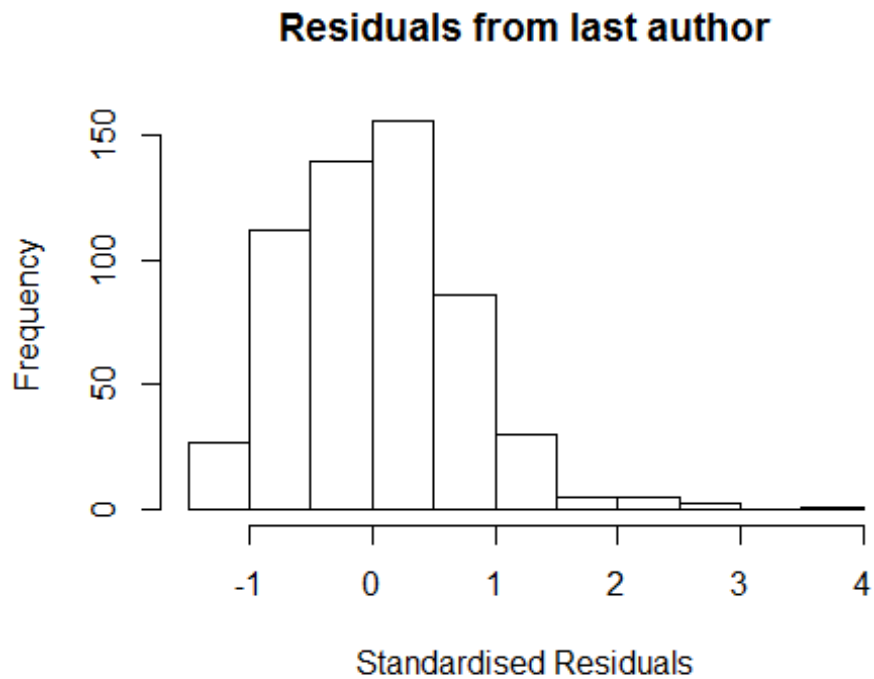
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

```



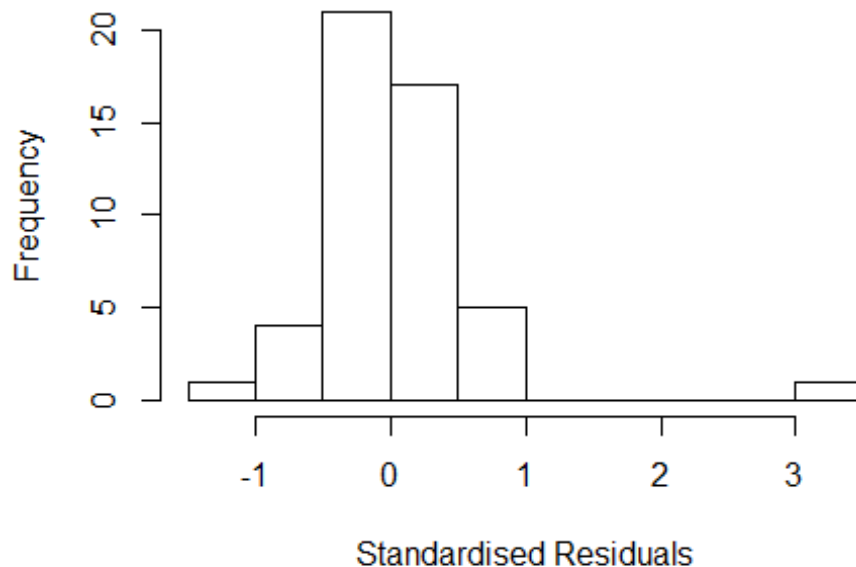
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in cov2cor(v): diag(.) had 0 or NA entries; non-finite result is
## doubtful
```



##	GVIF	Df	GVIF^(1/(2*Df))
## FirstAuthorFemale	NaN	1	NaN
## LastAuthorFemale	NaN	1	NaN
## UniqueAuthors	NaN	4	NaN
## Year	NaN	12	NaN

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
## 33 32044450086 3.239 2006      1701      2      3.317
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.28e+00 -2.61e-01  1.63e-15   2.88e-01  3.32e+00
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   -1.63e-15   0.00e+00   -Inf < 2e-16 ***
## FirstAuthorFemale1 -4.53e-01   2.46e-01  -1.84e+00  0.07589 .
## LastAuthorFemale1 -3.46e-01   1.97e-01  -1.75e+00  0.08948 .
## UniqueAuthors2     9.04e-02   1.71e-01   5.30e-01  0.60111
## UniqueAuthors3    -3.47e-02   2.81e-01  -1.20e-01  0.90264
## UniqueAuthors4     8.80e-01   2.07e-01   4.26e+00  0.00019 ***
## UniqueAuthors5     1.66e-01   1.07e-01   1.55e+00  0.13097
## Year2000           7.38e-01   0.00e+00    Inf < 2e-16 ***
## Year2001           7.27e-01   1.05e+00   6.90e-01  0.49412
## Year2003           1.25e+00   6.81e-01   1.84e+00  0.07577 .
```

```

## Year2004          1.62e+00    1.11e-08    1.46e+08    < 2e-16 ***
## Year2005          1.37e+00    3.42e-01    4.02e+00    0.00036 ***
## Year2006          6.31e-01    1.07e-01    5.90e+00    1.9e-06 ***
## Year2007          8.91e-01    1.50e-01    5.96e+00    1.6e-06 ***
## Year2008          1.57e+00    2.07e-01    7.58e+00    1.9e-08 ***
## Year2009          1.19e+00    2.64e-01    4.50e+00    9.6e-05 ***
## Year2010          7.92e-01    2.15e-01    3.69e+00    0.00089 ***
## Year2011          1.06e+00    1.22e-01    8.65e+00    1.2e-09 ***
## Year2012          1.67e+00    1.44e-01    1.15e+01    1.5e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.559, Adjusted R-squared:  0.295
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 11 is an outlier with |weight| = 0 ( < 0.002);
## 12 weights are ~= 1. The remaining 36 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.571  0.881  0.954   0.928  0.981   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          2.04e-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 4.011 1          2.003
## LastAuthorFemale  5.598 1          2.366
## Year              11.526 12          1.107

## [1] "List of 1 outliers with residuals above 2.5"
##   ScopusId  NLCS Year OneField Fields residuals
## 33 32044450086 3.239 2006    1701      2    3.368
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =

```

```

AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.3841 -0.2624 -0.0241  0.3809  3.3683
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   -2.05e-15   3.60e-08  0.00e+00  1.00000
## FirstAuthorFemale1 -5.71e-01   2.89e-01 -1.98e+00  0.05606 .
## LastAuthorFemale1 -2.38e-01   2.90e-01 -8.20e-01  0.41690
## Year2000         7.38e-01   3.60e-08  2.05e+07 < 2e-16 ***
## Year2001         8.17e-01   1.05e+00  7.80e-01  0.44003
## Year2003         1.23e+00   6.35e-01  1.95e+00  0.05995 .
## Year2004         1.62e+00   2.62e-08  6.19e+07 < 2e-16 ***
## Year2005         1.81e+00   2.02e-02  8.99e+01 < 2e-16 ***
## Year2006         6.80e-01   1.09e-01  6.22e+00  4.5e-07 ***
## Year2007         9.98e-01   1.98e-01  5.03e+00  1.6e-05 ***
## Year2008         1.99e+00   3.95e-01  5.04e+00  1.5e-05 ***
## Year2009         1.38e+00   3.17e-01  4.36e+00  0.00011 ***
## Year2010         8.54e-01   2.35e-01  3.64e+00  0.00090 ***
## Year2011         1.05e+00   1.44e-01  7.31e+00  1.8e-08 ***
## Year2012         1.66e+00   1.36e-01  1.22e+01  5.6e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.47,   Adjusted R-squared:  0.252
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 11 is an outlier with |weight| = 0 ( < 0.002);
## 11 weights are ~= 1. The remaining 37 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.506  0.876  0.950  0.914  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          2.04e-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

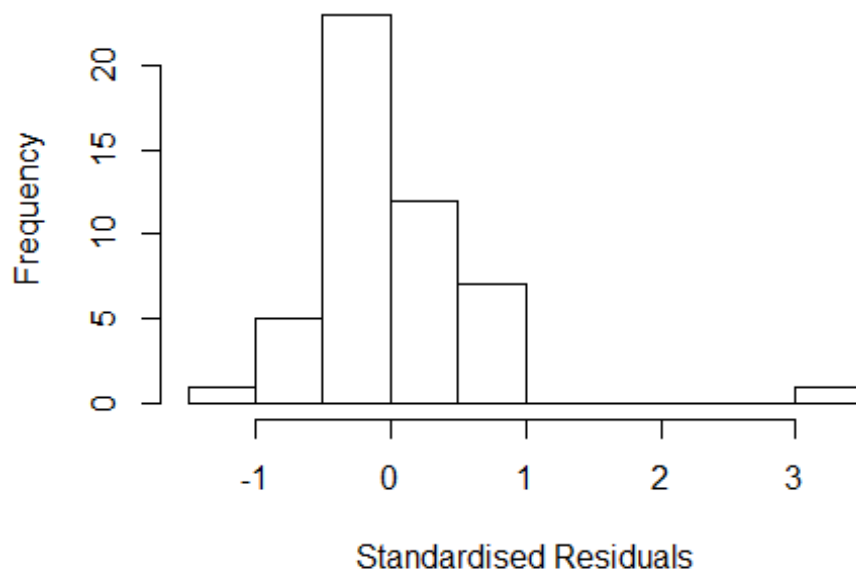
```

```
## "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

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

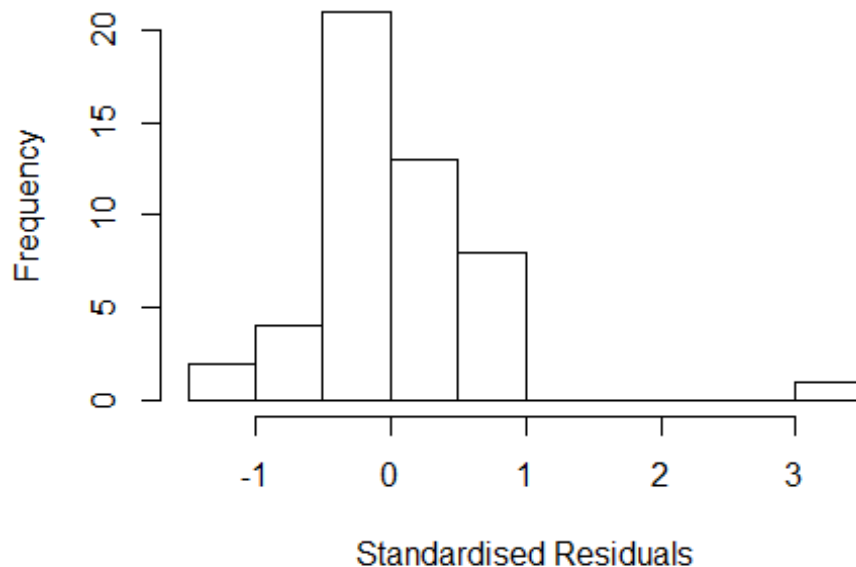
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```

Residuals from first and last author



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1          NaN
## Year              NaN 12         NaN
```

Residuals from first author



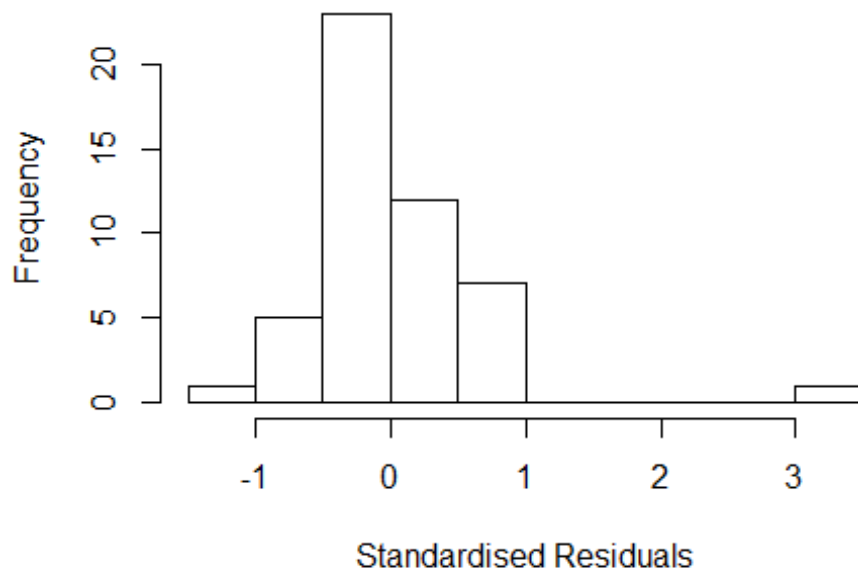
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 33 32044450086 3.239 2006      1701      2      3.368
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.40e+00 -2.99e-01 -2.22e-16  3.67e-01  3.36e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.33e-16   0.00e+00   -Inf    < 2e-16 ***
## FirstAuthorFemale1 -7.19e-01   2.15e-01   -3.34   0.00198 **
## Year2000        7.38e-01   0.00e+00    Inf    < 2e-16 ***
## Year2001        8.17e-01   1.17e+00    0.70   0.48861
## Year2003        1.23e+00   6.65e-01    1.86   0.07190 .
## Year2004        1.62e+00   0.00e+00    Inf    < 2e-16 ***
## Year2005        1.81e+00   2.02e-02   89.89    < 2e-16 ***
## Year2006        6.01e-01   9.78e-02    6.15   4.9e-07 ***
## Year2007        9.49e-01   1.84e-01    5.17   9.6e-06 ***
## Year2008        1.99e+00   4.03e-01    4.94   1.9e-05 ***
## Year2009        1.40e+00   3.02e-01    4.64   4.8e-05 ***
## Year2010        8.63e-01   2.36e-01    3.66   0.00082 ***
```

```

## Year2011          1.03e+00  1.67e-01   6.19  4.3e-07 ***
## Year2012          1.63e+00  1.53e-01  10.65  1.6e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.469, Adjusted R-squared:  0.272
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 11 is an outlier with |weight| = 0 ( < 0.002);
## 8 weights are ~= 1. The remaining 40 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.452  0.875  0.957  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          2.04e-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 1.977 1          1.406
## Year             1.977 12          1.029

```

Residuals from last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 33 32044450086 3.239 2006      1701      2      3.368
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + 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 -3.62e-01  3.18e-15   3.63e-01   2.98e+00
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   -3.18e-15   4.69e-08   0.00e+00  1.00000
## LastAuthorFemale1 -5.21e-01   2.66e-01  -1.96e+00  0.05791 .
## Year2000        7.38e-01   4.69e-08  1.57e+07 < 2e-16 ***
## Year2001        8.18e-01   8.35e-01   9.80e-01  0.33421
## Year2003        1.23e+00   5.69e-01   2.17e+00  0.03697 *
## Year2004        1.62e+00   4.70e-08  3.44e+07 < 2e-16 ***
## Year2005        1.81e+00   2.02e-02  8.99e+01 < 2e-16 ***
## Year2006        7.77e-01   9.86e-02   7.88e+00  2.9e-09 ***
## Year2007        1.01e+00   1.82e-01   5.57e+00  2.8e-06 ***
## Year2008        1.99e+00   3.77e-01   5.29e+00  6.7e-06 ***
## Year2009        1.27e+00   2.96e-01   4.28e+00  0.00014 ***
## Year2010        7.24e-01   2.56e-01   2.83e+00  0.00765 **
```

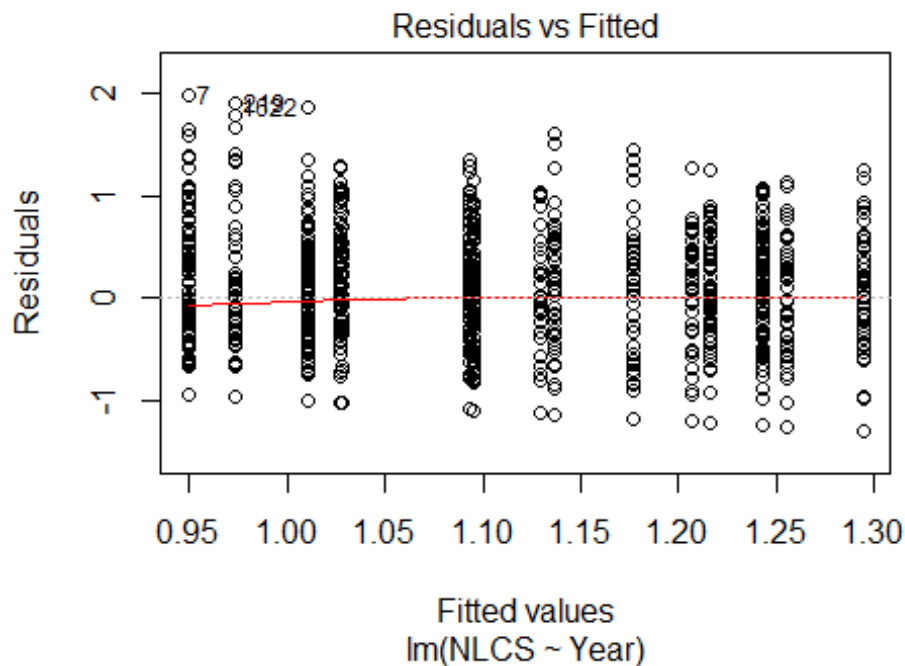


```

## Year2011          1.04e+00  1.40e-01  7.45e+00  1.0e-08 ***
## Year2012          1.57e+00  2.03e-01  7.70e+00  4.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.653
## Multiple R-squared:  0.404, Adjusted R-squared:  0.182
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 41 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0024 0.9110 0.9600 0.9140 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      2.04e-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: 49"
## [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
##   84   83   85   65   87   67   67   69   79   73   69  121  132  147  117
## 2011 2012
##  131  142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   52   59   32   46   32   45   41   52   49   39   73   82   99   71
## 2011 2012
##   82   96
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   44   49   25   38   30   35   34   46   42   30   61   74   86   62

```

```
## 2011 2012
## 74 76
## [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 = 0.84, df = 1, p-value = 0.4

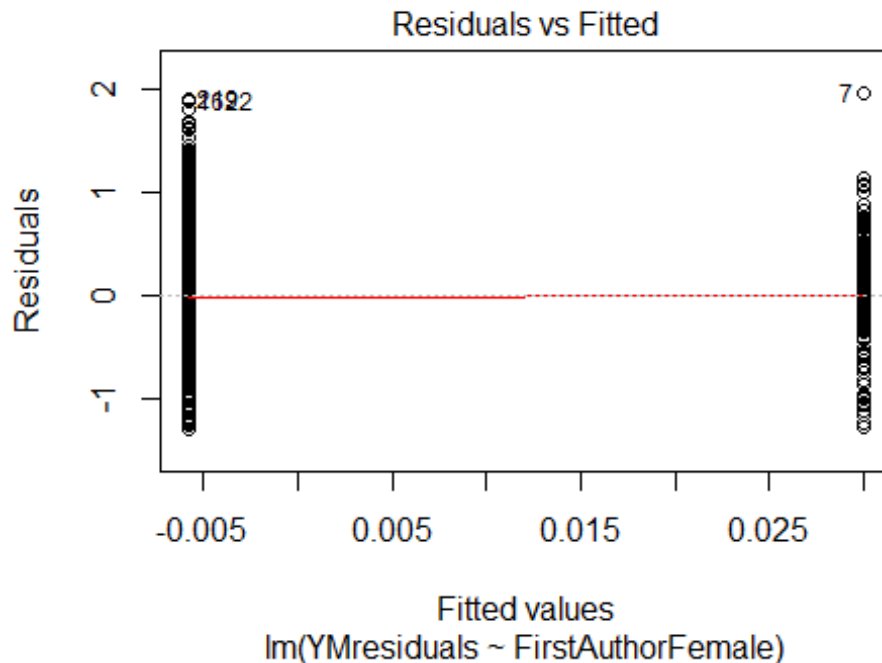
## [1] "Female first author team size 2018 geometric mean: 1.41421356237309"
## [1] "Male first author team size 2018 geometric mean: 2.7374762132844"

## 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 = 0.2
```

```
## 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: 2.28942848510666"

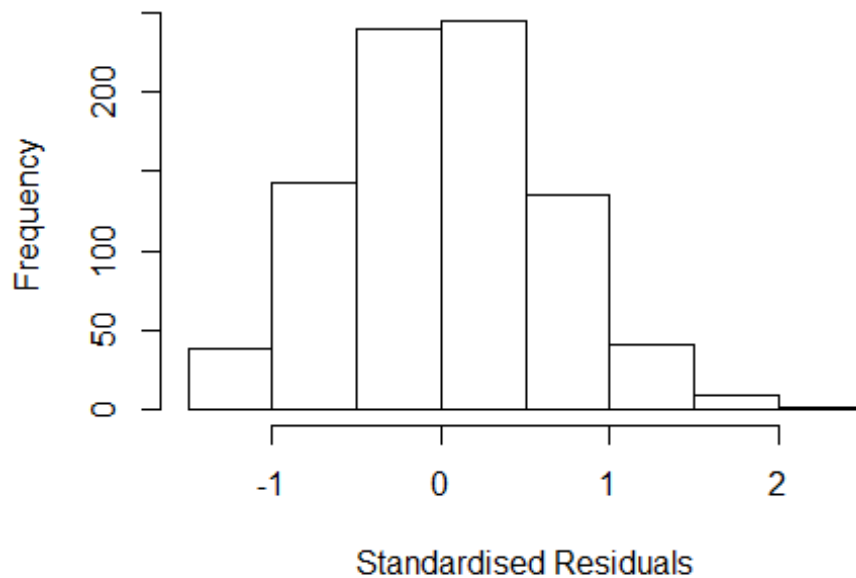
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12, 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	1.409	1	1.187
LastAuthorFemale	1.405	1	1.185
UniqueAuthors	1.407	4	1.044
Year	1.558	16	1.014

Residuals from first and last author and team size



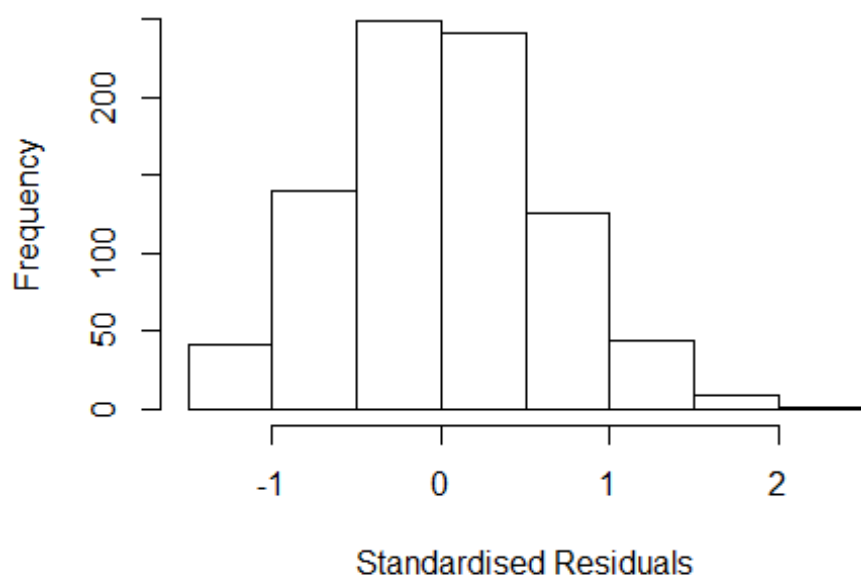
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3244 -0.4331 0.0123 0.4324 2.0679
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7515 0.1445 5.20 2.5e-07 ***
## FirstAuthorFemale1 0.0160 0.0650 0.25 0.8052
## LastAuthorFemale1 0.1072 0.0654 1.64 0.1016
## UniqueAuthors2 0.0996 0.0533 1.87 0.0621 .
## UniqueAuthors3 0.1997 0.0659 3.03 0.0025 **
## UniqueAuthors4 0.1686 0.1023 1.65 0.0998 .
## UniqueAuthors5 0.2461 0.1178 2.09 0.0371 *
## Year1997 0.0261 0.1722 0.15 0.8797
## Year1998 0.1214 0.1810 0.67 0.5026
## Year1999 0.2420 0.1927 1.26 0.2094
```

```

## Year2000          0.3591      0.1735      2.07      0.0388 *
## Year2001          0.2678      0.1858      1.44      0.1499
## Year2002          0.4733      0.1833      2.58      0.0100 **
## Year2003          0.3191      0.1878      1.70      0.0897 .
## Year2004          0.1652      0.1626      1.02      0.3101
## Year2005          0.1934      0.1836      1.05      0.2926
## Year2006          0.3449      0.1930      1.79      0.0743 .
## Year2007          0.3139      0.1640      1.91      0.0560 .
## Year2008          0.1158      0.1663      0.70      0.4866
## Year2009          0.2604      0.1589      1.64      0.1016
## Year2010          0.1003      0.1707      0.59      0.5568
## Year2011          0.3241      0.1618      2.00      0.0455 *
## Year2012          0.2126      0.1660      1.28      0.2007
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.636
## Multiple R-squared:  0.0567, Adjusted R-squared:  0.0316
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.269  0.871  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      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 1.413 1      1.189
## LastAuthorFemale  1.413 1      1.189
## Year              1.128 16      1.004

```

Residuals from first and last author



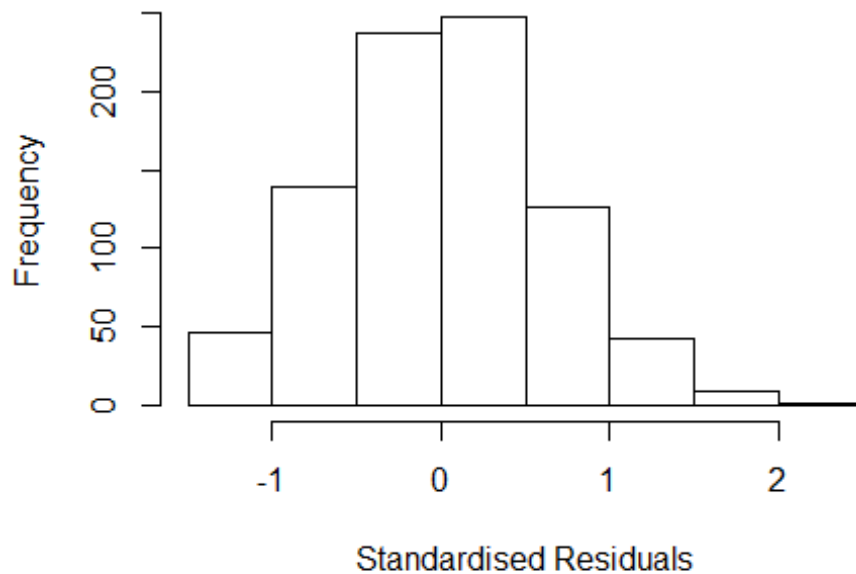
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32516 -0.43142 -0.00878 0.43230 2.12685
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7964 0.1377 5.78 1e-08 ***
## FirstAuthorFemale1 0.0118 0.0656 0.18 0.8579
## LastAuthorFemale1 0.1053 0.0661 1.59 0.1115
## Year1997 0.0377 0.1660 0.23 0.8202
## Year1998 0.1538 0.1767 0.87 0.3844
## Year1999 0.2882 0.1915 1.50 0.1327
## Year2000 0.3904 0.1691 2.31 0.0212 *
## Year2001 0.3015 0.1809 1.67 0.0959 .
## Year2002 0.5288 0.1772 2.98 0.0029 **
## Year2003 0.3551 0.1803 1.97 0.0493 *
## Year2004 0.1943 0.1578 1.23 0.2187
## Year2005 0.2390 0.1759 1.36 0.1748
```

```

## Year2006          0.3687      0.1930      1.91      0.0564 .
## Year2007          0.3588      0.1570      2.28      0.0226 *
## Year2008          0.1802      0.1603      1.12      0.2614
## Year2009          0.3216      0.1520      2.12      0.0346 *
## Year2010          0.1413      0.1647      0.86      0.3912
## Year2011          0.3880      0.1564      2.48      0.0133 *
## Year2012          0.2764      0.1589      1.74      0.0825 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0428, Adjusted R-squared:  0.0222
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 774 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.239  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      1.17e-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 1.077 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
## -1.33104 -0.43547 0.00232 0.43860 2.07541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8016 0.1385 5.79 1e-08 ***
## FirstAuthorFemale1 0.0579 0.0574 1.01 0.313
## Year1997 0.0397 0.1669 0.24 0.812
## Year1998 0.1492 0.1785 0.84 0.404
## Year1999 0.2911 0.1932 1.51 0.132
## Year2000 0.3957 0.1704 2.32 0.020 *
## Year2001 0.3074 0.1816 1.69 0.091 .
## Year2002 0.5294 0.1779 2.98 0.003 **
## Year2003 0.3556 0.1819 1.96 0.051 .
## Year2004 0.2136 0.1581 1.35 0.177
## Year2005 0.2446 0.1763 1.39 0.166
## Year2006 0.3648 0.1949 1.87 0.062 .
```

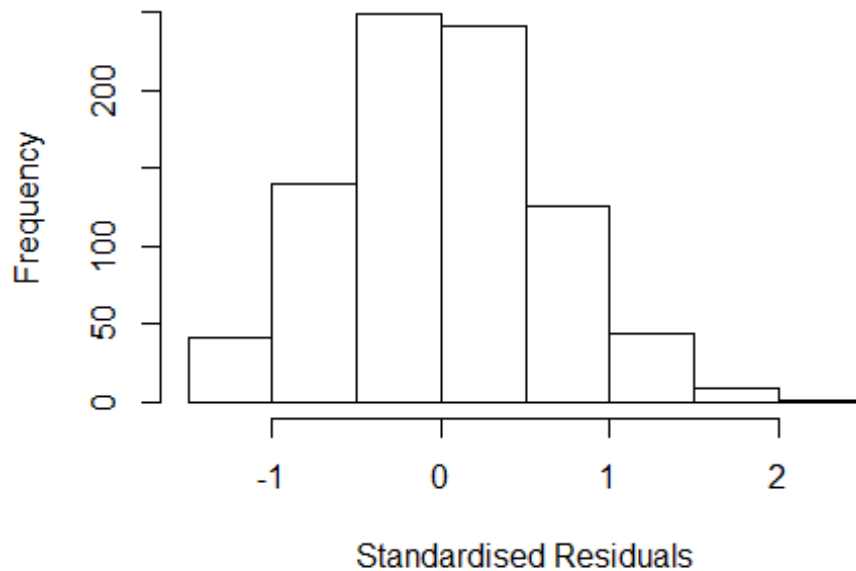


```

## Year2007          0.3624      0.1580      2.29      0.022 *
## Year2008          0.1861      0.1612      1.15      0.249
## Year2009          0.3305      0.1523      2.17      0.030 *
## Year2010          0.1483      0.1648      0.90      0.368
## Year2011          0.3973      0.1565      2.54      0.011 *
## Year2012          0.2811      0.1594      1.76      0.078 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0398, Adjusted R-squared:  0.0202
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.263  0.866  0.947  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      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 1.072 1      1.035
## Year              1.072 16      1.002

```

Residuals from last author



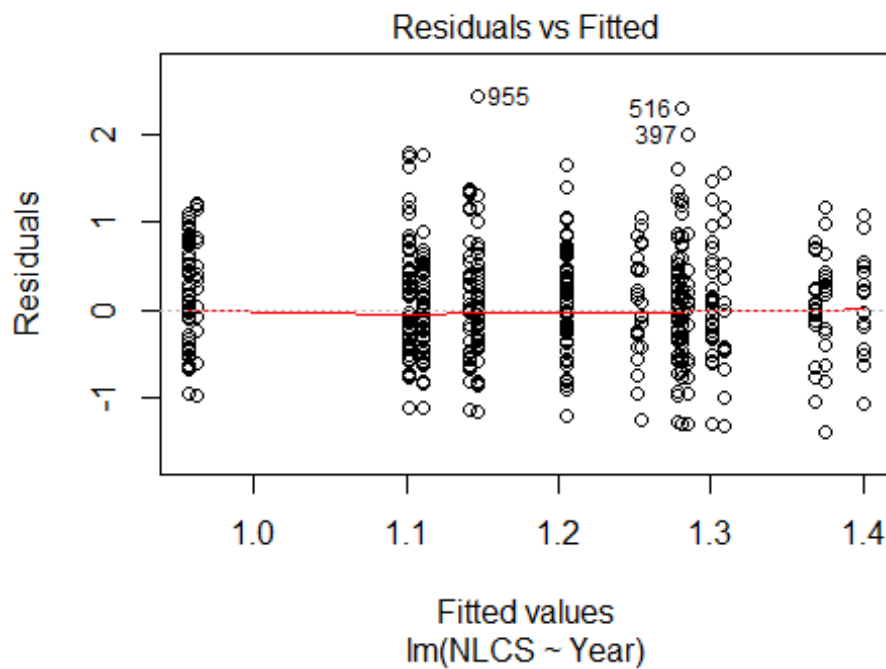
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3267 -0.4339 -0.0097 0.4339 2.1385
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7965 0.1379 5.78 1.1e-08 ***
## LastAuthorFemale1 0.1105 0.0577 1.92 0.0558 .
## Year1997 0.0381 0.1661 0.23 0.8187
## Year1998 0.1542 0.1766 0.87 0.3829
## Year1999 0.2889 0.1916 1.51 0.1319
## Year2000 0.3912 0.1689 2.32 0.0208 *
## Year2001 0.3017 0.1810 1.67 0.0960 .
## Year2002 0.5303 0.1770 3.00 0.0028 **
## Year2003 0.3565 0.1802 1.98 0.0482 *
## Year2004 0.1954 0.1578 1.24 0.2160
## Year2005 0.2391 0.1761 1.36 0.1750
## Year2006 0.3702 0.1922 1.93 0.0544 .
```

```

## Year2007          0.3602      0.1566      2.30      0.0217 *
## Year2008          0.1807      0.1604      1.13      0.2603
## Year2009          0.3236      0.1515      2.14      0.0330 *
## Year2010          0.1427      0.1643      0.87      0.3854
## Year2011          0.3893      0.1563      2.49      0.0129 *
## Year2012          0.2776      0.1587      1.75      0.0805 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.634
## Multiple R-squared:  0.0429, Adjusted R-squared:  0.0234
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 775 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.233  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      1.17e-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: 852"
## [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
##   42   47   46   43   43   43   42   48   58   44   72   71   95   95   82
## 2011 2012
##  103  109
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   15   14   15   15   19   15   17   33   32   43   35   61   51   53
## 2011 2012

```

```
## 72 72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 11 14 9 12 19 11 14 29 28 37 32 51 46 45
## 2011 2012
## 60 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 = 19, df = 16, p-value = 0.3
```



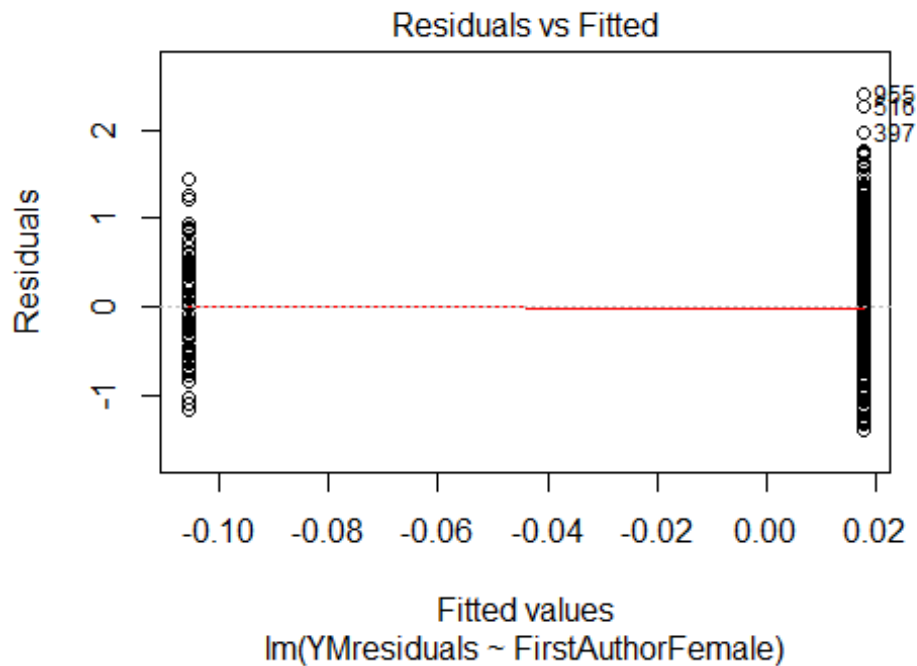
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 3.87010914244508"
## [1] "Male first author team size 2018 geometric mean: 2.21578758281976"

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

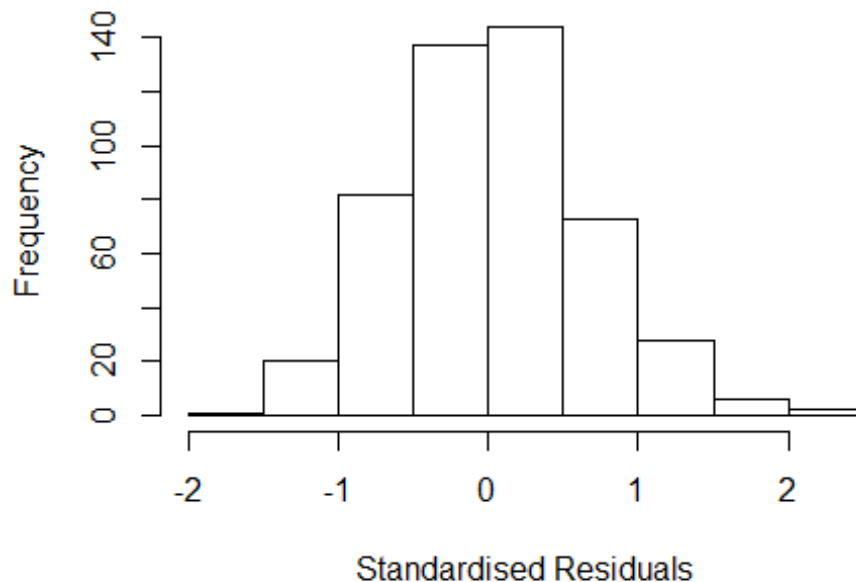
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.6207413942089"
## [1] "Male last author team size 2018 geometric mean: 2.39545453877627"

## 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 = 0.6
## 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 1.279  1      1.131
## LastAuthorFemale  1.301  1      1.141
## UniqueAuthors    1.858  4      1.081
## Year              2.161 16      1.024
```

Residuals from first and last author and team size



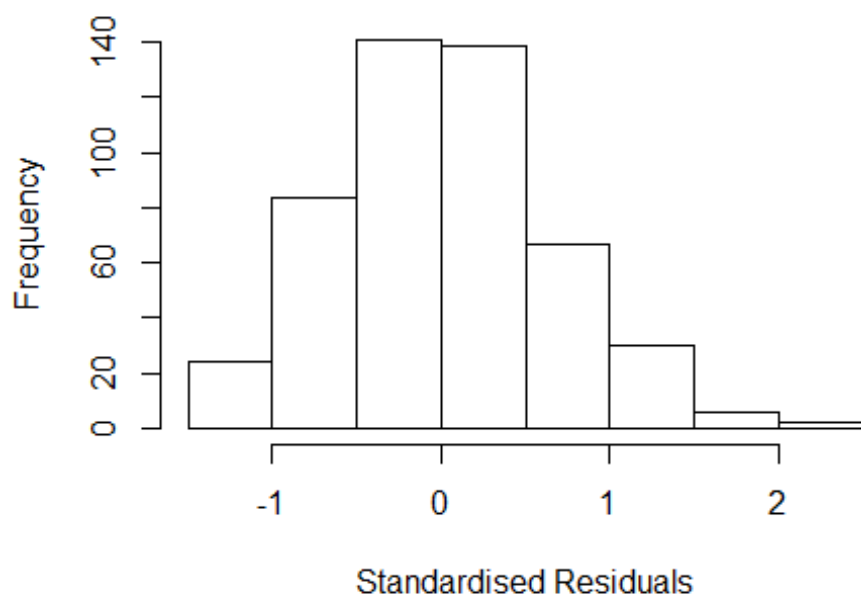
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5721 -0.4297 0.0166 0.4253 2.2892
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19618 0.26784 4.47 1.0e-05 ***
## FirstAuthorFemale1 -0.07224 0.08651 -0.84 0.40409
## LastAuthorFemale1 -0.05219 0.09835 -0.53 0.59589
## UniqueAuthors2 0.18845 0.07875 2.39 0.01710 *
## UniqueAuthors3 0.38495 0.08789 4.38 1.5e-05 ***
## UniqueAuthors4 0.46783 0.12117 3.86 0.00013 ***
## UniqueAuthors5 0.45285 0.09501 4.77 2.5e-06 ***
## Year1997 -0.29760 0.35921 -0.83 0.40781
## Year1998 -0.00887 0.35254 -0.03 0.97993
## Year1999 0.12560 0.31138 0.40 0.68687
```

```

## Year2000      0.01538      0.31231      0.05      0.96075
## Year2001     -0.32539      0.33881     -0.96      0.33736
## Year2002     -0.47056      0.31217     -1.51      0.13238
## Year2003     -0.11039      0.30909     -0.36      0.72113
## Year2004     -0.09779      0.30331     -0.32      0.74729
## Year2005     -0.08998      0.28793     -0.31      0.75480
## Year2006     -0.09191      0.27822     -0.33      0.74129
## Year2007     -0.17011      0.28773     -0.59      0.55467
## Year2008     -0.51544      0.27473     -1.88      0.06125 .
## Year2009     -0.29633      0.27656     -1.07      0.28450
## Year2010     -0.30491      0.27465     -1.11      0.26748
## Year2011     -0.37446      0.27651     -1.35      0.17631
## Year2012     -0.22699      0.27554     -0.82      0.41048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.634
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0666
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.165  0.873  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      2.03e-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 1.241 1      1.114
## LastAuthorFemale  1.210 1      1.100
## Year              1.294 16      1.008

```

Residuals from first and last author



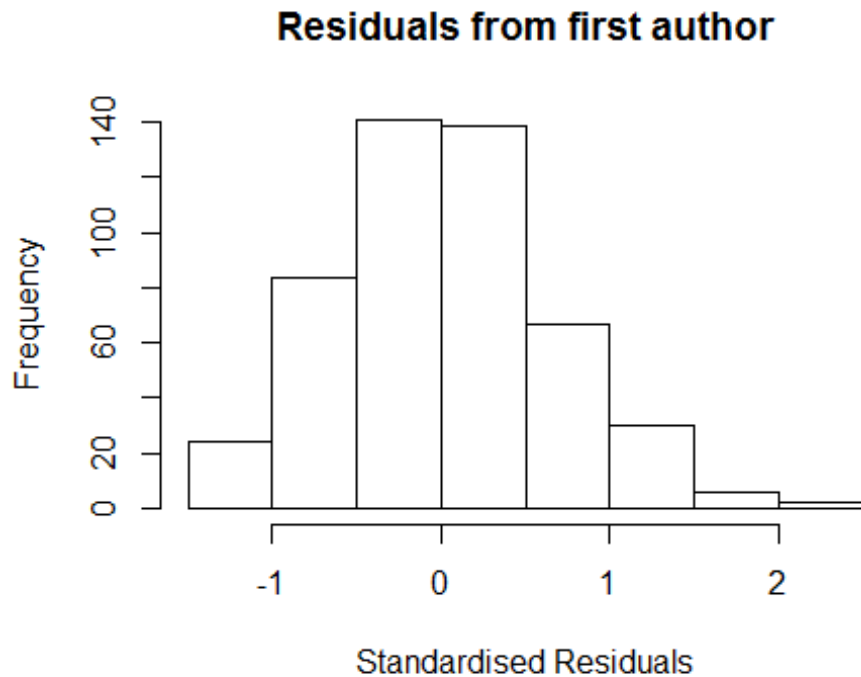
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31523 -0.41319 -0.00992 0.41973 2.44021
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.313865 0.234590 5.60 3.6e-08 ***
## FirstAuthorFemale1 -0.065784 0.087760 -0.75 0.45
## LastAuthorFemale1 0.000572 0.096167 0.01 1.00
## Year1997 -0.347294 0.336568 -1.03 0.30
## Year1998 -0.067213 0.343444 -0.20 0.84
## Year1999 0.151150 0.288165 0.52 0.60
## Year2000 0.037742 0.291250 0.13 0.90
## Year2001 -0.335830 0.335533 -1.00 0.32
## Year2002 -0.320536 0.290849 -1.10 0.27
## Year2003 -0.004679 0.295082 -0.02 0.99
## Year2004 -0.047345 0.276383 -0.17 0.86
## Year2005 0.001362 0.265506 0.01 1.00
```



```

## Year2006      -0.007984    0.253354   -0.03    0.97
## Year2007      -0.149633    0.268479   -0.56    0.58
## Year2008      -0.366832    0.250783   -1.46    0.14
## Year2009      -0.180074    0.249178   -0.72    0.47
## Year2010      -0.201596    0.247751   -0.81    0.42
## Year2011      -0.281941    0.251852   -1.12    0.26
## Year2012      -0.065469    0.247931   -0.26    0.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.641
## Multiple R-squared:  0.0449, Adjusted R-squared:  0.00863
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.116  0.873   0.954   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      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 1.145 1      1.070
## Year              1.145 16      1.004

```

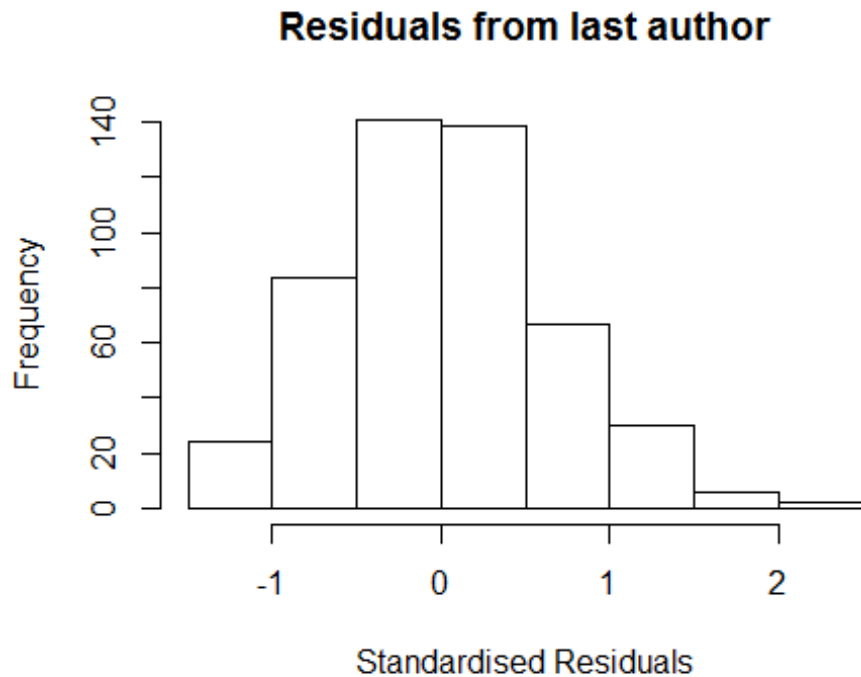


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3154 -0.4131 -0.0097 0.4197 2.4403
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31414 0.23460 5.60 3.6e-08 ***
## FirstAuthorFemale1 -0.06558 0.08445 -0.78 0.44
## Year1997 -0.34757 0.33675 -1.03 0.30
## Year1998 -0.06783 0.34356 -0.20 0.84
## Year1999 0.15101 0.28868 0.52 0.60
## Year2000 0.03751 0.29130 0.13 0.90
## Year2001 -0.33616 0.33545 -1.00 0.32
## Year2002 -0.32066 0.29095 -1.10 0.27
## Year2003 -0.00504 0.29505 -0.02 0.99
## Year2004 -0.04782 0.27617 -0.17 0.86
## Year2005 0.00122 0.26502 0.00 1.00
## Year2006 -0.00829 0.25315 -0.03 0.97
```

```

## Year2007          -0.14998    0.26862   -0.56    0.58
## Year2008          -0.36712    0.25099   -1.46    0.14
## Year2009          -0.18046    0.24933   -0.72    0.47
## Year2010          -0.20188    0.24792   -0.81    0.42
## Year2011          -0.28244    0.25198   -1.12    0.26
## Year2012          -0.06566    0.24785   -0.26    0.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.045, Adjusted R-squared:  0.0108
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.114  0.872  0.954  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      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 1.119 1          1.058
## Year            1.119 16          1.004

```



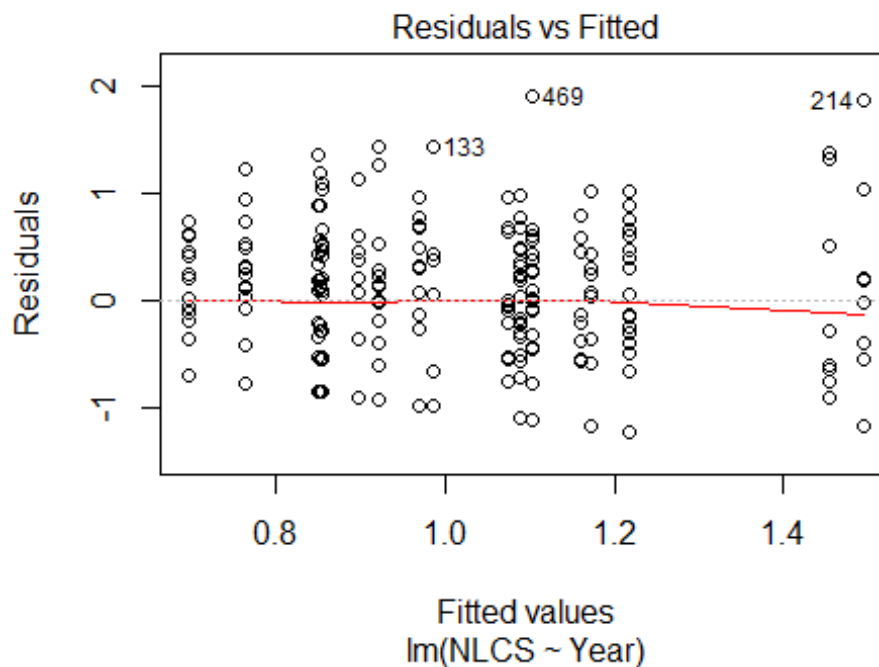
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31602 -0.41798 -0.00551 0.42169 2.44650
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30512 0.23023 5.67 2.5e-08 ***
## LastAuthorFemale1 -0.02005 0.09242 -0.22 0.83
## Year1997 -0.34086 0.33067 -1.03 0.30
## Year1998 -0.05881 0.34062 -0.17 0.86
## Year1999 0.15242 0.28802 0.53 0.60
## Year2000 0.04653 0.28778 0.16 0.87
## Year2001 -0.34966 0.33070 -1.06 0.29
## Year2002 -0.31329 0.28638 -1.09 0.27
## Year2003 -0.00118 0.29302 0.00 1.00
## Year2004 -0.04649 0.27358 -0.17 0.87
## Year2005 0.01090 0.26153 0.04 0.97
## Year2006 -0.00331 0.24991 -0.01 0.99
```

```

## Year2007          -0.15325      0.26446    -0.58      0.56
## Year2008          -0.36913      0.24648    -1.50      0.13
## Year2009          -0.17762      0.24521    -0.72      0.47
## Year2010          -0.20410      0.24343    -0.84      0.40
## Year2011          -0.27761      0.24781    -1.12      0.26
## Year2012          -0.06338      0.24401    -0.26      0.80
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.0441, Adjusted R-squared:  0.00984
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.112  0.871  0.954  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      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 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
##   26   32   13   19   18   22   21   22   16   24   36   33   34   28   24
## 2011 2012
##   50   30
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   11   20    8   12    8    7   12   15    9   14   24   18   20    9   13
## 2011 2012

```

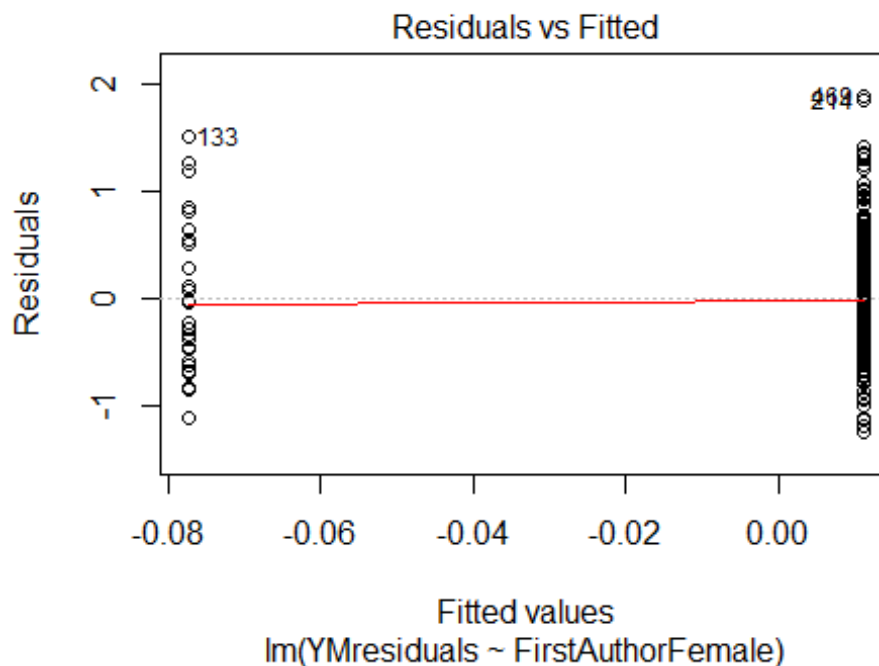
```
## 26 19
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 17 5 12 8 6 10 12 7 13 23 17 19 7 12
## 2011 2012
## 21 17
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 12, df = 16, p-value = 0.8
```



```
##
## 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: 2.03393700979443"
## [1] "Male first author team size 2018 geometric mean: 2.88506202716261"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

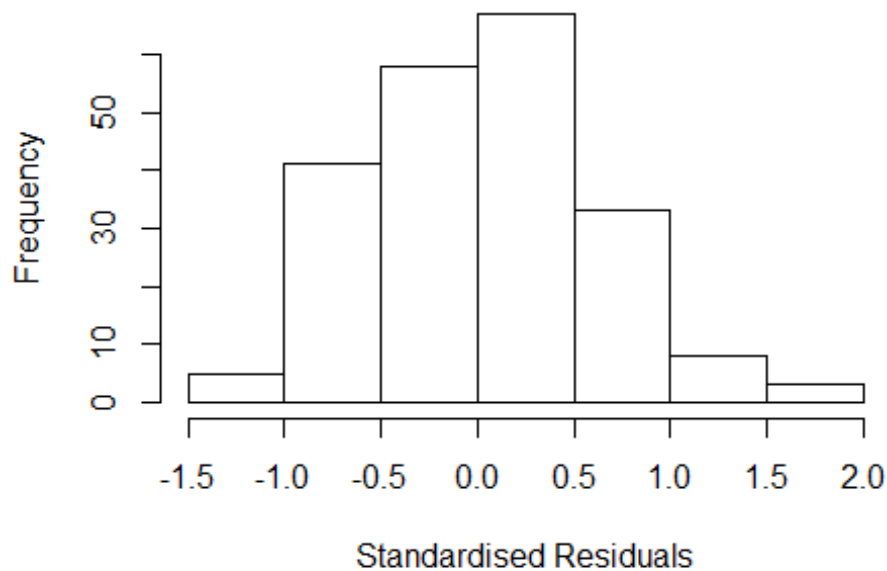
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 46, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.14610795432031"
## [1] "Male last author team size 2018 geometric mean: 2.86973743839547"

## 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 = 50, 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 1.882  1      1.372
## LastAuthorFemale  1.940  1      1.393
## UniqueAuthors    3.258  4      1.159
## Year              4.777 16      1.050
```

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
## -1.2132 -0.4577 0.0129 0.3896 1.9141
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.458428 0.304078 1.51 0.1333
## FirstAuthorFemale1 -0.036876 0.162401 -0.23 0.8206
## LastAuthorFemale1 -0.224432 0.174898 -1.28 0.2010
## UniqueAuthors2 0.590550 0.117304 5.03 1.1e-06 ***
## UniqueAuthors3 0.625346 0.134863 4.64 6.5e-06 ***
## UniqueAuthors4 0.441345 0.143414 3.08 0.0024 **
## UniqueAuthors5 0.196722 0.137740 1.43 0.1549
## Year1997 -0.000746 0.327662 0.00 0.9982
## Year1998 0.947263 0.427333 2.22 0.0278 *
## Year1999 0.131564 0.382552 0.34 0.7313
```

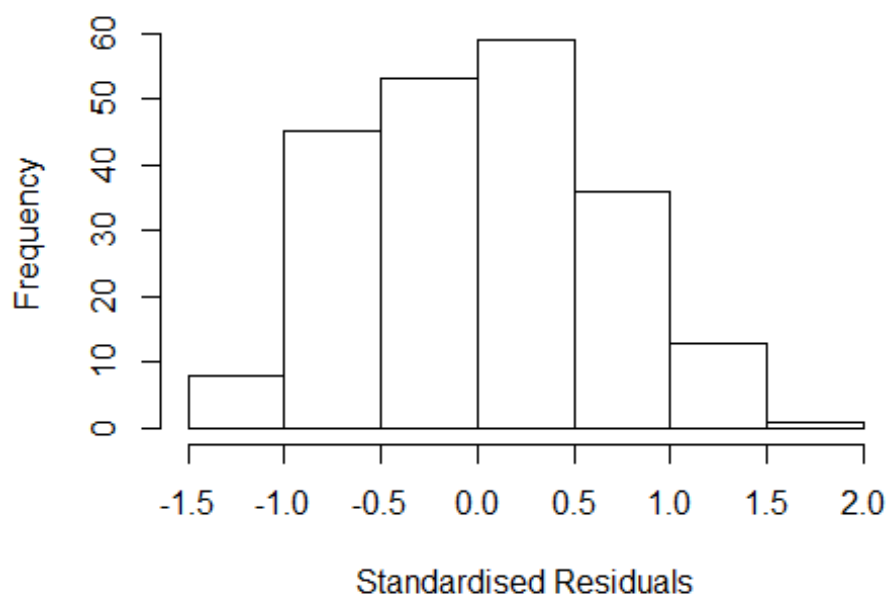


```

## Year2000      0.314612    0.363710    0.87    0.3881
## Year2001      0.277785    0.379521    0.73    0.4651
## Year2002      0.476679    0.354882    1.34    0.1808
## Year2003      0.308808    0.355637    0.87    0.3863
## Year2004      0.562845    0.546283    1.03    0.3042
## Year2005      0.129525    0.326243    0.40    0.6918
## Year2006      0.388653    0.315021    1.23    0.2188
## Year2007      0.100861    0.308791    0.33    0.7443
## Year2008      0.095893    0.319571    0.30    0.7644
## Year2009      0.281775    0.342442    0.82    0.4116
## Year2010     -0.076183    0.347879   -0.22    0.8269
## Year2011      0.274698    0.321343    0.85    0.3937
## Year2012      0.274403    0.337271    0.81    0.4169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.59
## Multiple R-squared:  0.243, Adjusted R-squared:  0.156
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 199 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.877  0.945  0.912  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 1.645 1      1.283
## LastAuthorFemale  1.645 1      1.283
## Year              1.726 16      1.017

```

Residuals from first and last author



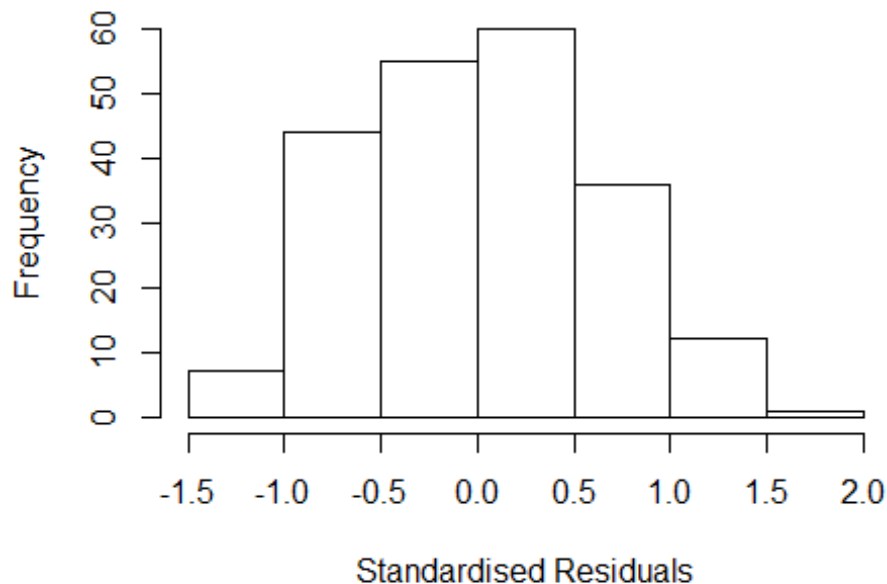
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1609 -0.4899 0.0401 0.4526 1.8459
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.83725 0.24911 3.36 0.00093 ***
## FirstAuthorFemale1 -0.02232 0.16582 -0.13 0.89307
## LastAuthorFemale1 -0.09753 0.17434 -0.56 0.57651
## Year1997 -0.08634 0.29644 -0.29 0.77117
## Year1998 0.59967 0.47500 1.26 0.20829
## Year1999 0.00209 0.32214 0.01 0.99484
## Year2000 0.35377 0.31417 1.13 0.26153
## Year2001 0.23574 0.44549 0.53 0.59728
## Year2002 0.29985 0.30265 0.99 0.32304
## Year2003 0.15067 0.33365 0.45 0.65207
## Year2004 0.67784 0.46906 1.45 0.15002
## Year2005 0.17529 0.34826 0.50 0.61529
```

```

## Year2006          0.24483    0.27029    0.91  0.36615
## Year2007          -0.08832    0.27787   -0.32  0.75093
## Year2008          -0.00874    0.29833   -0.03  0.97666
## Year2009           0.22078    0.33533    0.66  0.51106
## Year2010          -0.03940    0.32854   -0.12  0.90465
## Year2011           0.13917    0.28795    0.48  0.62942
## Year2012           0.32368    0.29819    1.09  0.27904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.683
## Multiple R-squared:  0.0773, Adjusted R-squared:  -0.00748
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.445  0.891  0.951  0.923  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 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.265 1      1.125
## Year              1.265 16      1.007

```

Residuals from first author



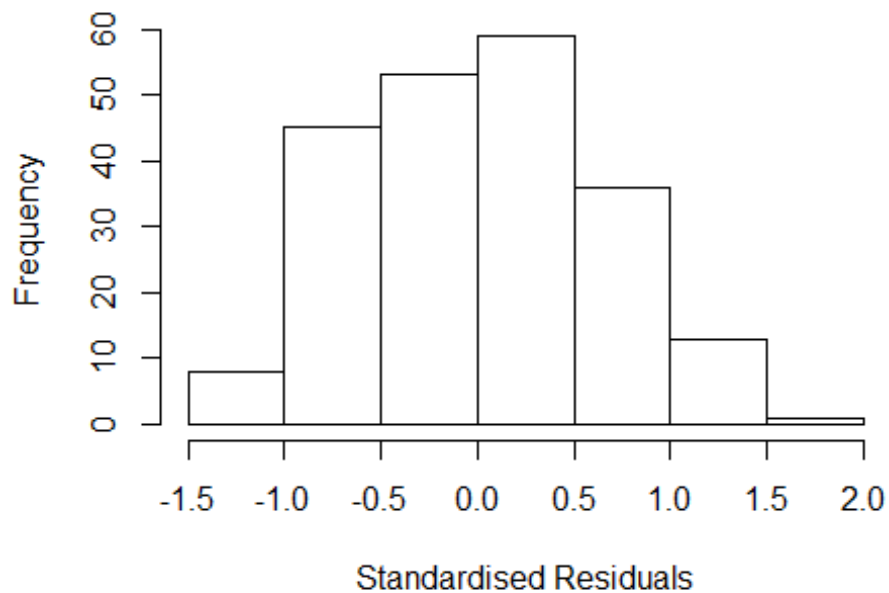
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1863 -0.4912 0.0414 0.4587 1.8497
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.824106 0.259057 3.18 0.0017 **
## FirstAuthorFemale1 -0.066971 0.150671 -0.44 0.6572
## Year1997 -0.076453 0.306555 -0.25 0.8033
## Year1998 0.602133 0.487638 1.23 0.2184
## Year1999 0.000841 0.330853 0.00 0.9980
## Year2000 0.335610 0.321245 1.04 0.2974
## Year2001 0.250383 0.456305 0.55 0.5838
## Year2002 0.300371 0.310941 0.97 0.3352
## Year2003 0.163994 0.338228 0.48 0.6283
## Year2004 0.687241 0.484293 1.42 0.1575
## Year2005 0.171291 0.355010 0.48 0.6300
## Year2006 0.242314 0.282198 0.86 0.3916
```

```

## Year2007          -0.078854    0.287652   -0.27    0.7843
## Year2008          -0.000149    0.305136    0.00    0.9996
## Year2009           0.243615    0.340704    0.72    0.4754
## Year2010          -0.026688    0.335357   -0.08    0.9367
## Year2011           0.145471    0.295678    0.49    0.6233
## Year2012           0.326845    0.307440    1.06    0.2890
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.691
## Multiple R-squared:  0.0755, Adjusted R-squared:  -0.00428
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 204 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.453  0.895   0.953   0.925   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 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.283 1          1.133
## Year              1.283 16          1.008

```

Residuals from last author



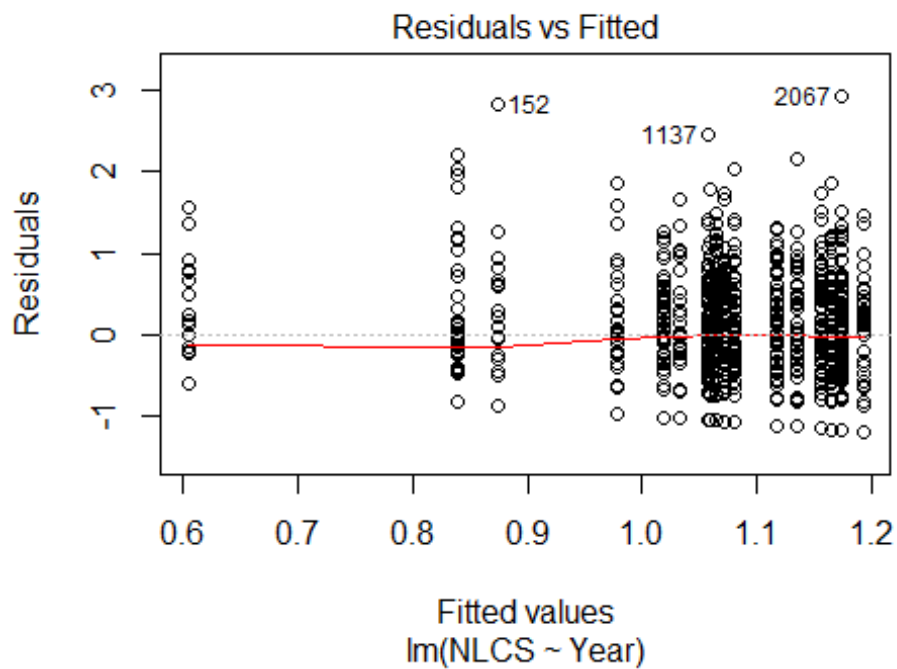
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1582 -0.4982 0.0403 0.4558 1.8473
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.83580 0.25001 3.34 0.00099 ***
## LastAuthorFemale1 -0.10674 0.15571 -0.69 0.49383
## Year1997 -0.08835 0.29255 -0.30 0.76297
## Year1998 0.59657 0.47465 1.26 0.21029
## Year1999 0.00296 0.32226 0.01 0.99267
## Year2000 0.35575 0.31596 1.13 0.26157
## Year2001 0.23639 0.44299 0.53 0.59421
## Year2002 0.30079 0.30372 0.99 0.32321
## Year2003 0.14834 0.33051 0.45 0.65406
## Year2004 0.67793 0.46994 1.44 0.15072
## Year2005 0.17743 0.34845 0.51 0.61119
## Year2006 0.24782 0.27316 0.91 0.36539
```

```

## Year2007          -0.08768      0.27778    -0.32  0.75261
## Year2008          -0.00914      0.29764    -0.03  0.97553
## Year2009           0.21406      0.32297     0.66  0.50824
## Year2010          -0.04218      0.32520    -0.13  0.89694
## Year2011           0.13871      0.28609     0.48  0.62831
## Year2012           0.32239      0.29507     1.09  0.27591
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.678
## Multiple R-squared:  0.0771, Adjusted R-squared:  -0.00257
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 202 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.439  0.890  0.951  0.922  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] "Sample size for the above analysis: 215"
## [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
##   65   67   90   68   56   60   70   70   84  118   92  129  136  131  138
## 2011 2012
##  124  144
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   28   29   55   32   33   32   45   39   54   70   60   78   91   74   83
## 2011 2012

```

```
## 71 86
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 24 49 27 28 28 34 34 44 68 49 63 78 62 71
## 2011 2012
## 59 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 = 15, df = 16, p-value = 0.5
```

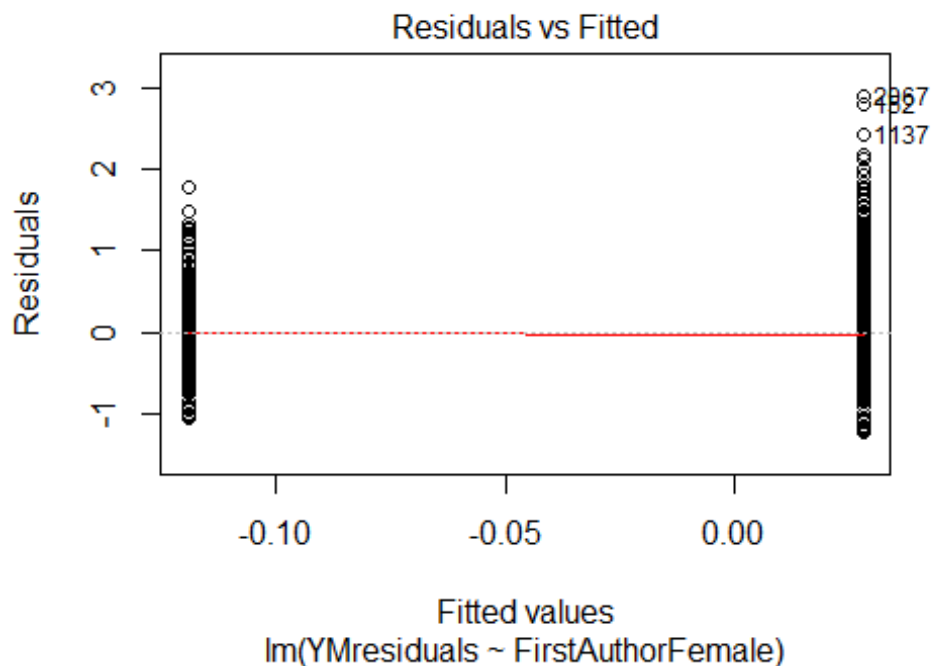


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
## [1] "Female first author team size 2018 geometric mean: 3.17372650520247"
## [1] "Male first author team size 2018 geometric mean: 2.25371602848289"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



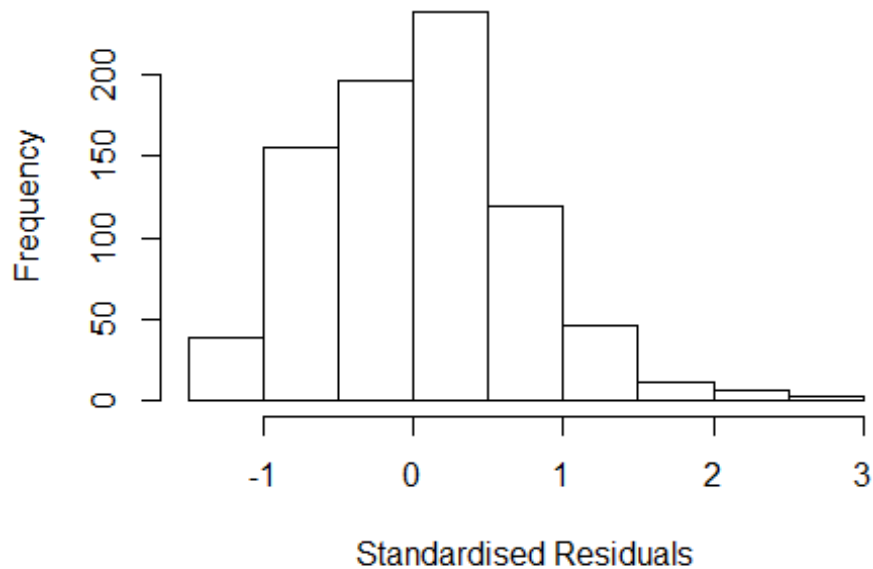
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.87422218668489"
## [1] "Male last author team size 2018 geometric mean: 2.31061830246676"

## 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 1.529  1      1.237
## LastAuthorFemale  1.564  1      1.251
## UniqueAuthors    1.456  4      1.048
## Year              1.610 16      1.015
```

Residuals from first and last author and team size



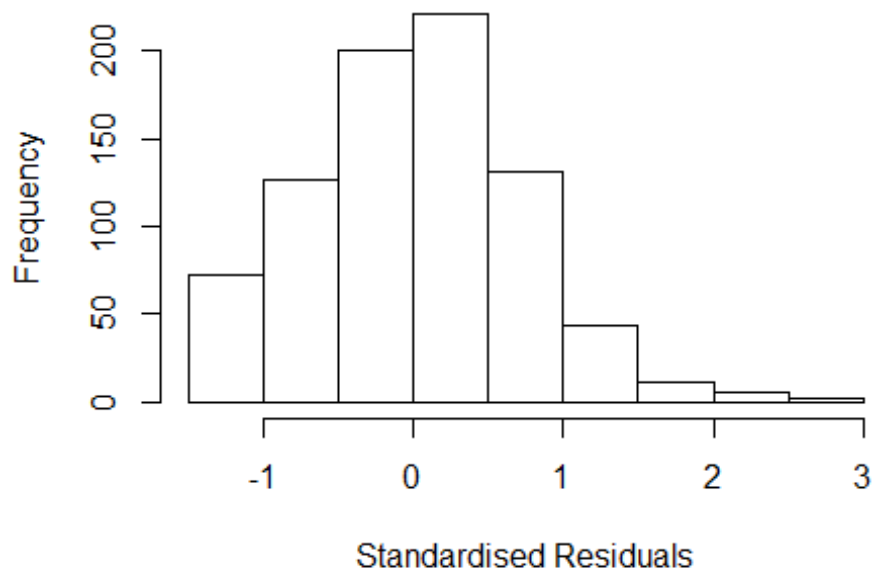
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 152   0031482189 3.717 1997    1705     4     2.933
## 2067 84857370722 4.100 2012    1705     3     2.744
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4741 -0.4535  0.0324  0.4475  2.9334
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.53061    0.12520   4.24 2.5e-05 ***
## FirstAuthorFemale1 -0.11954    0.06446  -1.85 0.06406 .
## LastAuthorFemale1  0.03036    0.06426   0.47 0.63673
## UniqueAuthors2    0.00773    0.06060   0.13 0.89859
## UniqueAuthors3    0.31150    0.06980   4.46 9.3e-06 ***
## UniqueAuthors4    0.34421    0.10512   3.27 0.00110 **
## UniqueAuthors5    0.40180    0.16279   2.47 0.01379 *
## Year1997         0.25299    0.23503   1.08 0.28207
## Year1998         0.17405    0.17266   1.01 0.31373
```

```

## Year1999      0.33464      0.20109      1.66  0.09648 .
## Year2000      0.38198      0.18947      2.02  0.04413 *
## Year2001      0.51362      0.17963      2.86  0.00436 **
## Year2002      0.56899      0.17894      3.18  0.00153 **
## Year2003      0.60956      0.16302      3.74  0.00020 ***
## Year2004      0.40332      0.15787      2.55  0.01081 *
## Year2005      0.44459      0.15284      2.91  0.00373 **
## Year2006      0.49289      0.17519      2.81  0.00502 **
## Year2007      0.44320      0.14619      3.03  0.00251 **
## Year2008      0.37633      0.14970      2.51  0.01214 *
## Year2009      0.45956      0.15191      3.03  0.00257 **
## Year2010      0.45839      0.15079      3.04  0.00244 **
## Year2011      0.56894      0.15413      3.69  0.00024 ***
## Year2012      0.51355      0.14596      3.52  0.00046 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.68
## Multiple R-squared:  0.0893, Adjusted R-squared:  0.0639
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 741 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0232 0.8560 0.9490 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      1.23e-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 1.398 1      1.182
## LastAuthorFemale 1.447 1      1.203
## Year      1.161 16      1.005

```

Residuals from first and last author



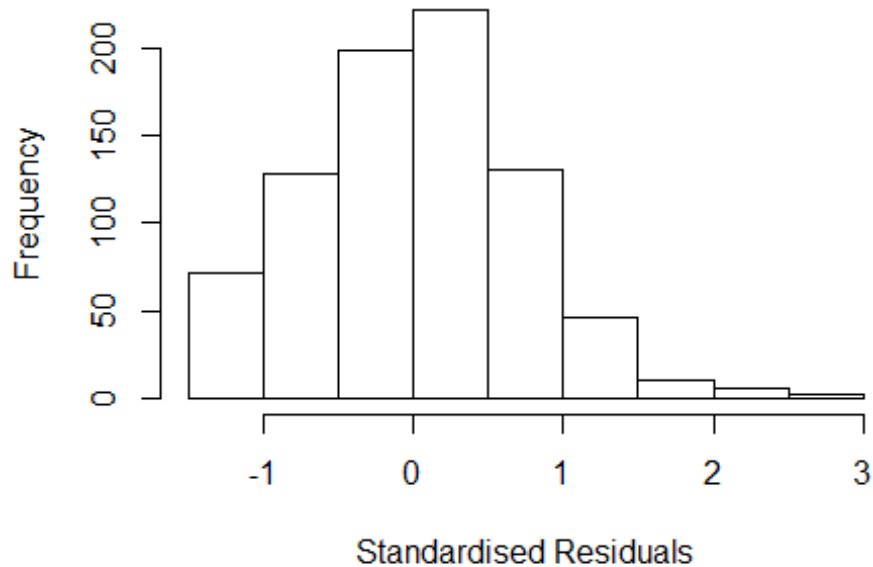
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 152   0031482189 3.717 1997    1705     4     2.850
## 2067 84857370722 4.100 2012    1705     3     2.913
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2456 -0.4762  0.0147  0.4660  2.9128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5877    0.1204   4.88 1.3e-06 ***
## FirstAuthorFemale1 -0.1390    0.0642  -2.17 0.03064 *
## LastAuthorFemale1  0.0247    0.0635   0.39 0.69754
## Year1997         0.2789    0.2186   1.28 0.20244
## Year1998         0.1392    0.1698   0.82 0.41237
## Year1999         0.2979    0.1934   1.54 0.12380
## Year2000         0.3875    0.1861   2.08 0.03764 *
## Year2001         0.5272    0.1818   2.90 0.00384 **
## Year2002         0.5975    0.1798   3.32 0.00093 ***
## Year2003         0.6212    0.1605   3.87 0.00012 ***
## Year2004         0.4733    0.1611   2.94 0.00339 **
```

```

## Year2005          0.5036      0.1514      3.33  0.00092 ***
## Year2006          0.5264      0.1749      3.01  0.00270 **
## Year2007          0.4817      0.1454      3.31  0.00096 ***
## Year2008          0.4445      0.1489      2.99  0.00291 **
## Year2009          0.5308      0.1476      3.60  0.00034 ***
## Year2010          0.5700      0.1469      3.88  0.00011 ***
## Year2011          0.6332      0.1516      4.18  3.3e-05 ***
## Year2012          0.5995      0.1430      4.19  3.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.702
## Multiple R-squared:  0.0479, Adjusted R-squared:  0.0263
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 741 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0465 0.8670 0.9460 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      1.23e-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 1.067 1      1.033
## Year      1.067 16      1.002

```

Residuals from first author



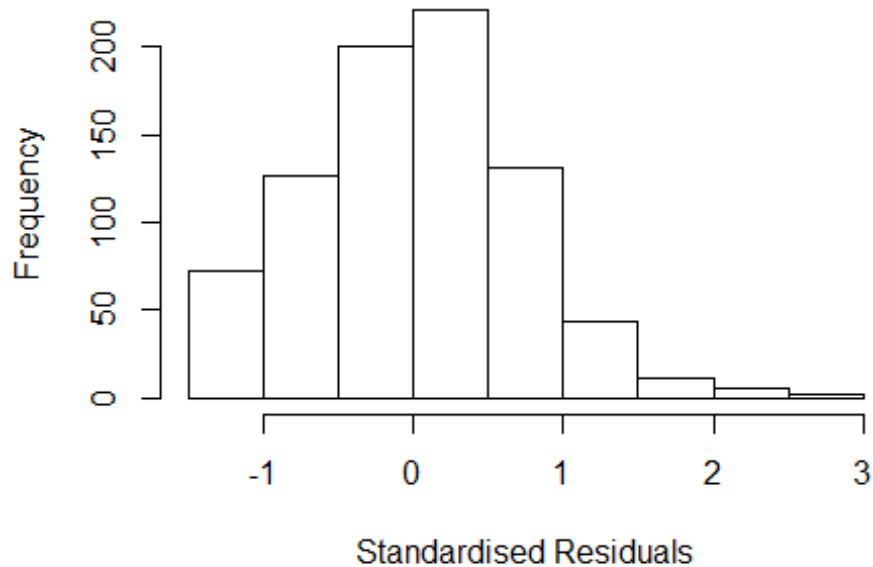
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 152   0031482189 3.717 1997    1705     4     2.850
## 2067 84857370722 4.100 2012    1705     3     2.913
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2249 -0.4682  0.0141  0.4621  2.9121
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5868    0.1204   4.87 1.3e-06 ***
## FirstAuthorFemale1 -0.1260    0.0562  -2.24 0.02518 *
## Year1997        0.2824    0.2190   1.29 0.19752
## Year1998        0.1407    0.1695   0.83 0.40673
## Year1999        0.3035    0.1933   1.57 0.11671
## Year2000        0.3901    0.1860   2.10 0.03627 *
## Year2001        0.5318    0.1812   2.93 0.00344 **
## Year2002        0.6004    0.1797   3.34 0.00087 ***
## Year2003        0.6278    0.1586   3.96 8.2e-05 ***
## Year2004        0.4759    0.1613   2.95 0.00327 **
## Year2005        0.5083    0.1506   3.38 0.00077 ***
```

```

## Year2006          0.5284      0.1748      3.02  0.00258 **
## Year2007          0.4836      0.1451      3.33  0.00090 ***
## Year2008          0.4468      0.1486      3.01  0.00272 **
## Year2009          0.5362      0.1461      3.67  0.00026 ***
## Year2010          0.5746      0.1461      3.93  9.1e-05 ***
## Year2011          0.6380      0.1507      4.23  2.6e-05 ***
## Year2012          0.6011      0.1429      4.21  2.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.0479, Adjusted R-squared:  0.0276
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 743 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0448 0.8670 0.9460 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      1.23e-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 1.108 1      1.053
## Year      1.108 16      1.003

```

Residuals from last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 152   0031482189 3.717 1997    1705     4     2.850
## 2067 84857370722 4.100 2012    1705     3     2.913
##
## Call:
## lmrob(formula = NLCS ~ 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 -0.4884  0.0256  0.4801  2.9382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.5818    0.1205   4.83 1.7e-06 ***
## LastAuthorFemale1 -0.0481    0.0559  -0.86 0.38901
## Year1997          0.2898    0.2183   1.33 0.18459
## Year1998          0.1258    0.1696   0.74 0.45838
## Year1999          0.2925    0.1945   1.50 0.13314
## Year2000          0.3804    0.1882   2.02 0.04355 *
## Year2001          0.5197    0.1826   2.85 0.00455 **
## Year2002          0.5916    0.1814   3.26 0.00116 **
## Year2003          0.6159    0.1603   3.84 0.00013 ***
## Year2004          0.4639    0.1617   2.87 0.00422 **
## Year2005          0.4996    0.1523   3.28 0.00108 **
```

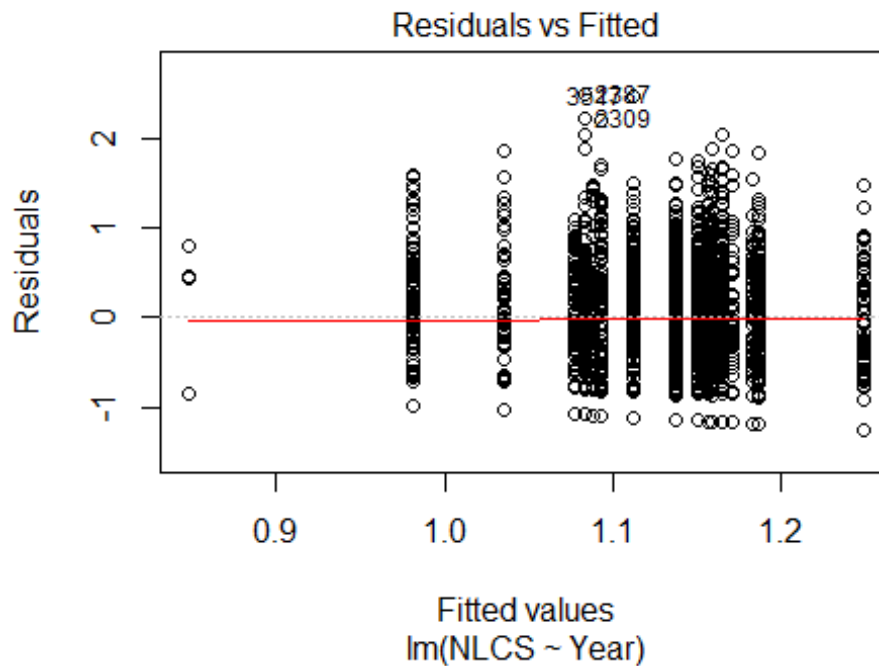


```

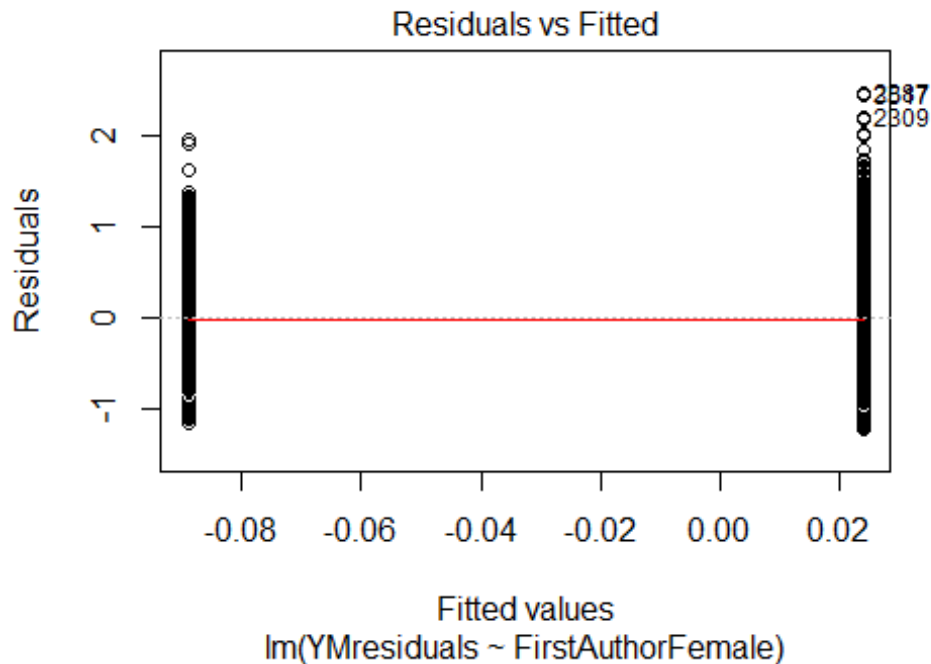
## Year2006          0.5089      0.1734      2.93  0.00344 **
## Year2007          0.4698      0.1456      3.23  0.00130 **
## Year2008          0.4393      0.1490      2.95  0.00328 **
## Year2009          0.5257      0.1480      3.55  0.00040 ***
## Year2010          0.5690      0.1475      3.86  0.00012 ***
## Year2011          0.6315      0.1511      4.18  3.3e-05 ***
## Year2012          0.5800      0.1429      4.06  5.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.705
## Multiple R-squared:  0.0436, Adjusted R-squared:  0.0232
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 742 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0434 0.8700 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      1.23e-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: 814"
## [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
## 200 12 156 169 182 243 180 144 194 208 294 304 368 441 400
## 2011 2012
## 447 431
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 5 60 74 58 92 104 73 94 117 179 170 224 243 251

```

```
## 2011 2012
## 264 264
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 63 5 52 64 51 82 87 59 76 100 145 143 190 203 220
## 2011 2012
## 227 224
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```

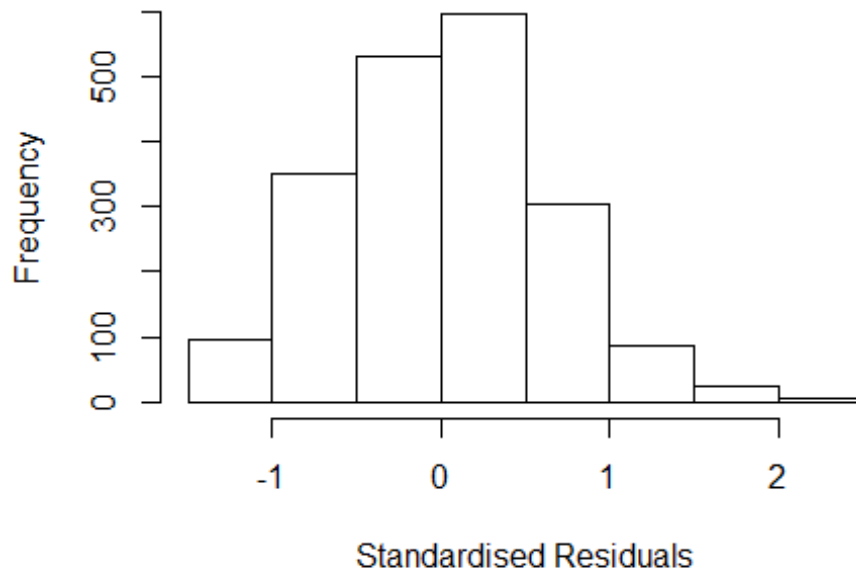


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.18493305112582"
## [1] "Male first author team size 2018 geometric mean: 2.92101557939732"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 990, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.79291189435518"
## [1] "Male last author team size 2018 geometric mean: 3.03795799418058"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 690, 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 1.316 1      1.147
## LastAuthorFemale  1.348 1      1.161
## UniqueAuthors    1.189 4      1.022
## Year              1.209 16     1.006
```

Residuals from first and last author and team size



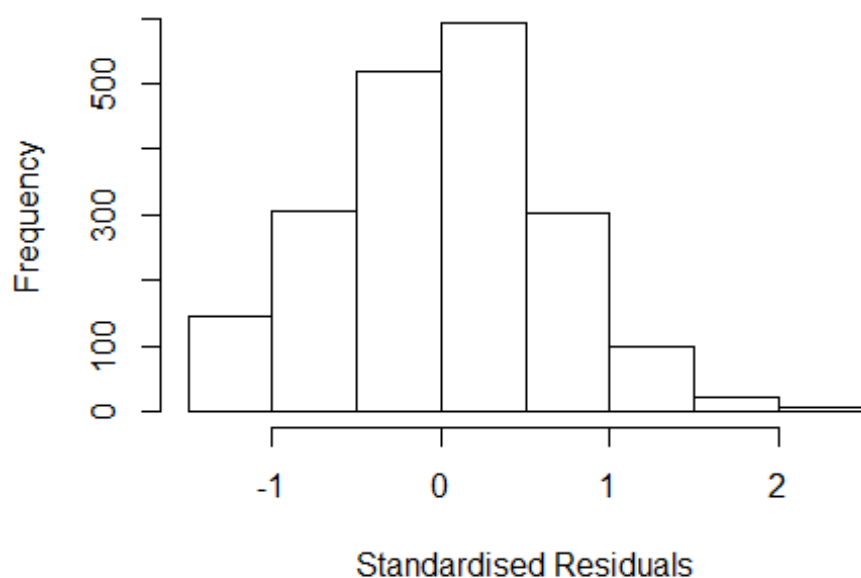
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3286 -0.4328 0.0194 0.4164 2.4630
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.91039 0.10099 9.01 < 2e-16 ***
## FirstAuthorFemale1 -0.04365 0.03909 -1.12 0.264
## LastAuthorFemale1 -0.06420 0.04201 -1.53 0.127
## UniqueAuthors2 0.25729 0.03809 6.76 1.9e-11 ***
## UniqueAuthors3 0.33565 0.04138 8.11 8.7e-16 ***
## UniqueAuthors4 0.35777 0.05183 6.90 6.9e-12 ***
## UniqueAuthors5 0.36705 0.06023 6.09 1.3e-09 ***
## Year1997 -0.18251 0.35037 -0.52 0.602
## Year1998 0.11055 0.14657 0.75 0.451
## Year1999 0.21215 0.12712 1.67 0.095 .
```

```

## Year2000      0.13624    0.12319    1.11    0.269
## Year2001      0.05120    0.13489    0.38    0.704
## Year2002     -0.00062    0.13552    0.00    0.996
## Year2003     -0.01072    0.13327   -0.08    0.936
## Year2004     -0.05076    0.13047   -0.39    0.697
## Year2005     -0.15484    0.12054   -1.28    0.199
## Year2006     -0.05570    0.11350   -0.49    0.624
## Year2007      0.10118    0.11483    0.88    0.378
## Year2008     -0.00848    0.11159   -0.08    0.939
## Year2009     -0.03001    0.10873   -0.28    0.783
## Year2010     -0.01414    0.11121   -0.13    0.899
## Year2011      0.03922    0.10882    0.36    0.719
## Year2012     -0.00659    0.11046   -0.06    0.952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.636
## Multiple R-squared:  0.0655, Adjusted R-squared:  0.055
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 166 weights are ~= 1. The remaining 1825 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.100  0.872  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      5.02e-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 1.262 1      1.123
## LastAuthorFemale  1.267 1      1.125
## Year              1.050 16      1.002

```

Residuals from first and last author



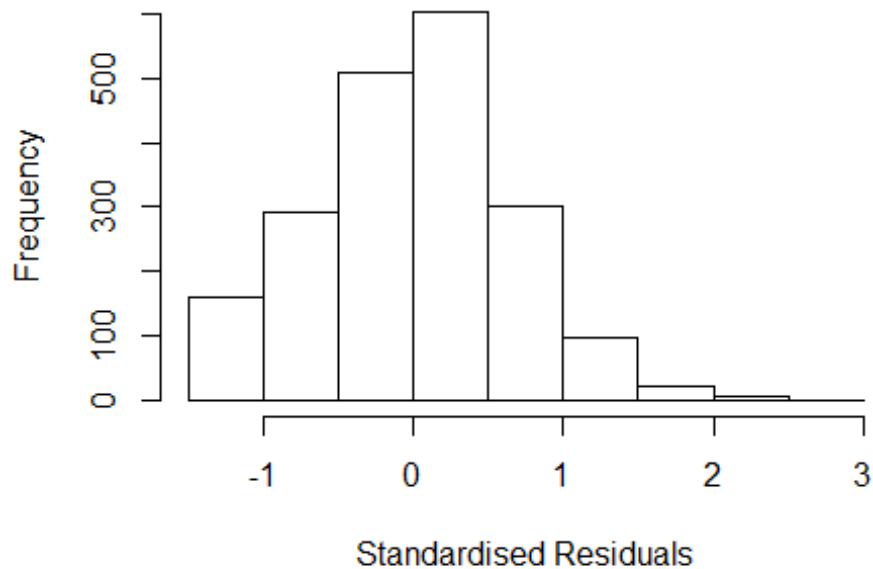
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3104 -0.4489 0.0282 0.4271 2.4905
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0353 0.0990 10.46 <2e-16 ***
## FirstAuthorFemale1 -0.0492 0.0392 -1.26 0.209
## LastAuthorFemale1 -0.1033 0.0419 -2.46 0.014 *
## Year1997 -0.1568 0.4028 -0.39 0.697
## Year1998 0.1357 0.1447 0.94 0.349
## Year1999 0.2751 0.1279 2.15 0.032 *
## Year2000 0.1605 0.1263 1.27 0.204
## Year2001 0.0990 0.1353 0.73 0.464
## Year2002 0.0674 0.1337 0.50 0.614
## Year2003 0.0672 0.1323 0.51 0.611
## Year2004 0.0151 0.1276 0.12 0.906
## Year2005 -0.0482 0.1202 -0.40 0.689
```

```

## Year2006          0.0509      0.1119      0.46      0.649
## Year2007          0.1899      0.1137      1.67      0.095 .
## Year2008          0.1067      0.1095      0.97      0.330
## Year2009          0.0483      0.1072      0.45      0.653
## Year2010          0.0993      0.1099      0.90      0.366
## Year2011          0.1413      0.1071      1.32      0.187
## Year2012          0.1138      0.1085      1.05      0.295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0158, Adjusted R-squared:  0.00683
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1834 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.872  0.951  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      5.02e-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 1.025 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
## -1.2983 -0.4517 0.0307 0.4320 2.5003
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0329 0.0999 10.34 <2e-16 ***
## FirstAuthorFemale1 -0.0936 0.0355 -2.63 0.0085 **
## Year1997 -0.1546 0.4031 -0.38 0.7013
## Year1998 0.1265 0.1441 0.88 0.3803
## Year1999 0.2654 0.1283 2.07 0.0387 *
## Year2000 0.1566 0.1277 1.23 0.2200
## Year2001 0.0930 0.1354 0.69 0.4924
## Year2002 0.0589 0.1339 0.44 0.6603
## Year2003 0.0583 0.1329 0.44 0.6609
## Year2004 0.0151 0.1281 0.12 0.9062
## Year2005 -0.0513 0.1214 -0.42 0.6727
## Year2006 0.0428 0.1130 0.38 0.7049
```

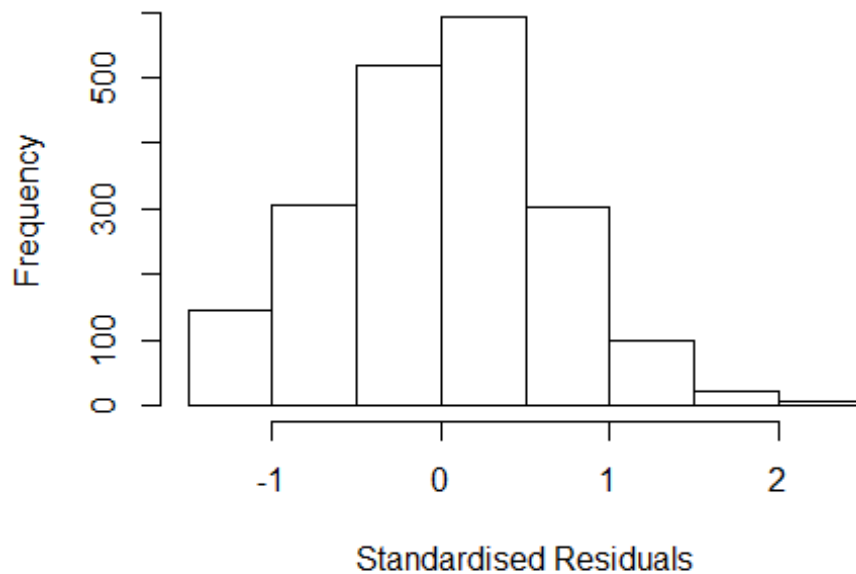


```

## Year2007          0.1793      0.1145      1.57      0.1173
## Year2008          0.0988      0.1102      0.90      0.3702
## Year2009          0.0408      0.1082      0.38      0.7065
## Year2010          0.0927      0.1106      0.84      0.4020
## Year2011          0.1305      0.1079      1.21      0.2267
## Year2012          0.0993      0.1093      0.91      0.3635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.652
## Multiple R-squared:  0.0125, Adjusted R-squared:  0.00401
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1828 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.872  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      5.02e-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 1.03 1      1.015
## Year      1.03 16      1.001

```

Residuals from last author



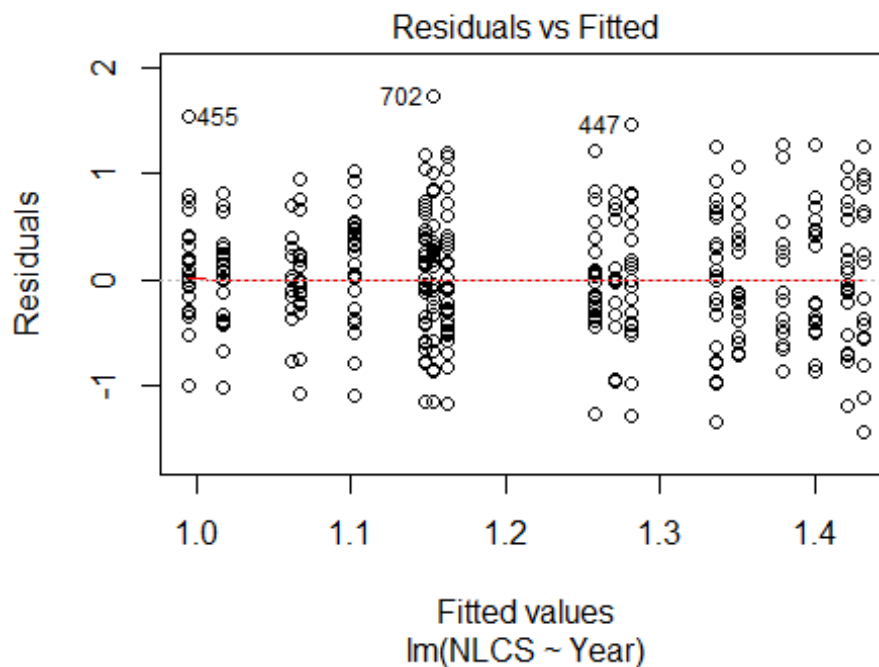
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3030 -0.4444 0.0281 0.4286 2.4973
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0297 0.0989 10.42 < 2e-16 ***
## LastAuthorFemale1 -0.1266 0.0379 -3.34 0.00085 ***
## Year1997 -0.1513 0.4028 -0.38 0.70723
## Year1998 0.1389 0.1452 0.96 0.33864
## Year1999 0.2733 0.1279 2.14 0.03280 *
## Year2000 0.1582 0.1258 1.26 0.20863
## Year2001 0.0980 0.1350 0.73 0.46789
## Year2002 0.0682 0.1340 0.51 0.61084
## Year2003 0.0651 0.1320 0.49 0.62214
## Year2004 0.0146 0.1278 0.11 0.90879
## Year2005 -0.0489 0.1202 -0.41 0.68395
## Year2006 0.0513 0.1120 0.46 0.64693
```

```

## Year2007          0.1890      0.1136      1.66  0.09653 .
## Year2008          0.1022      0.1092      0.94  0.34970
## Year2009          0.0469      0.1071      0.44  0.66137
## Year2010          0.0962      0.1098      0.88  0.38113
## Year2011          0.1413      0.1070      1.32  0.18660
## Year2012          0.1142      0.1084      1.05  0.29243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.00659
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1830 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.872  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      5.02e-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: 1991"
## [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
##   27   59   38   43   51   38   29   34   24   29   36   51   31   49   49
## 2011 2012
##   41   60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   35   19   18   28   19   17   21   11   20   27   26   13   23   30
## 2011 2012

```

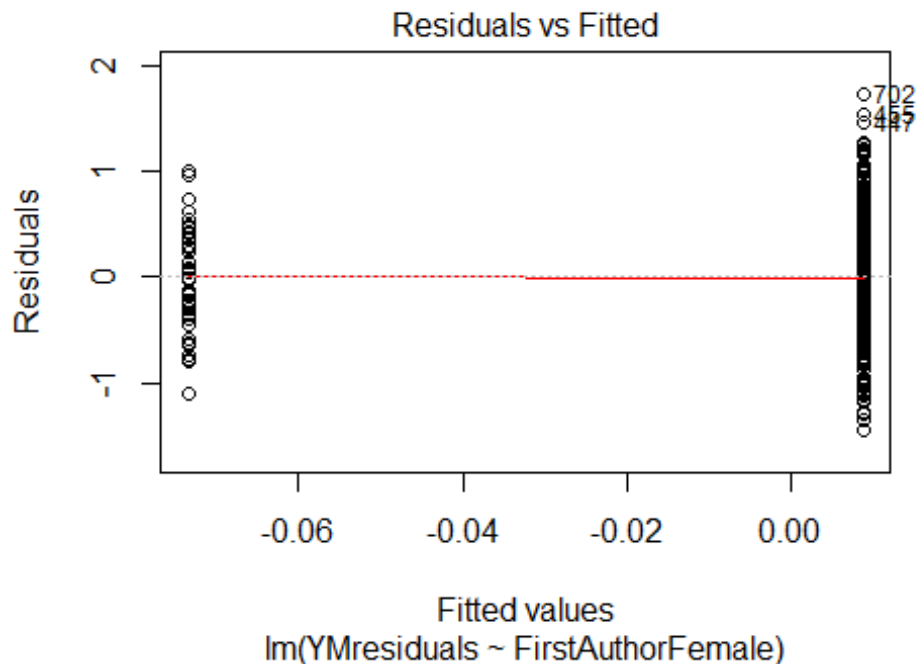
```
## 25 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 25 13 11 21 18 13 17 9 16 24 20 11 21 26
## 2011 2012
## 24 32
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```



```
##
## 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: 3.22370979547063"
## [1] "Male first author team size 2018 geometric mean: 2.72096526527247"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

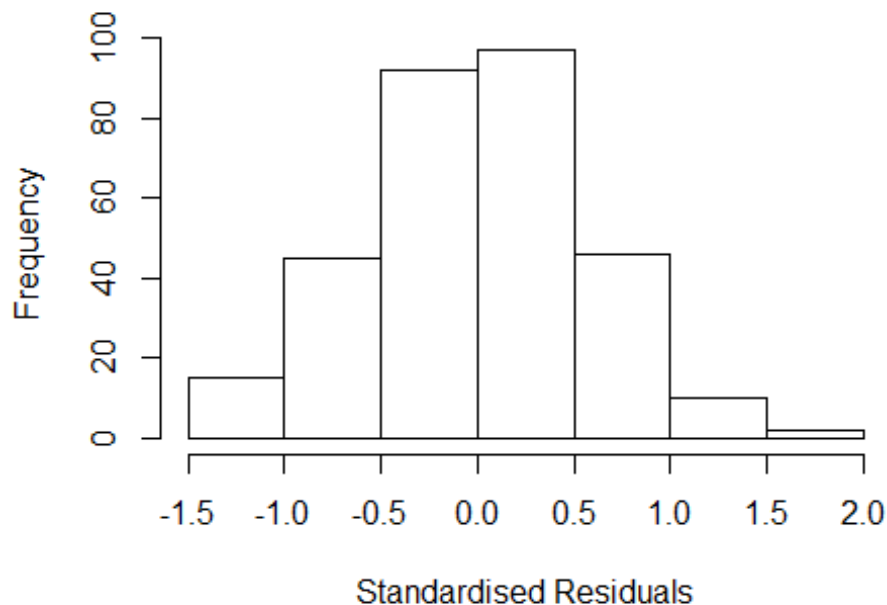
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.46410161513775"
## [1] "Male last author team size 2018 geometric mean: 2.75239669709557"

## 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 = 0.4
## 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 1.960  1      1.400
## LastAuthorFemale  2.088  1      1.445
## UniqueAuthors    2.211  4      1.104
## Year              4.052 16      1.045
```

Residuals from first and last author and team size



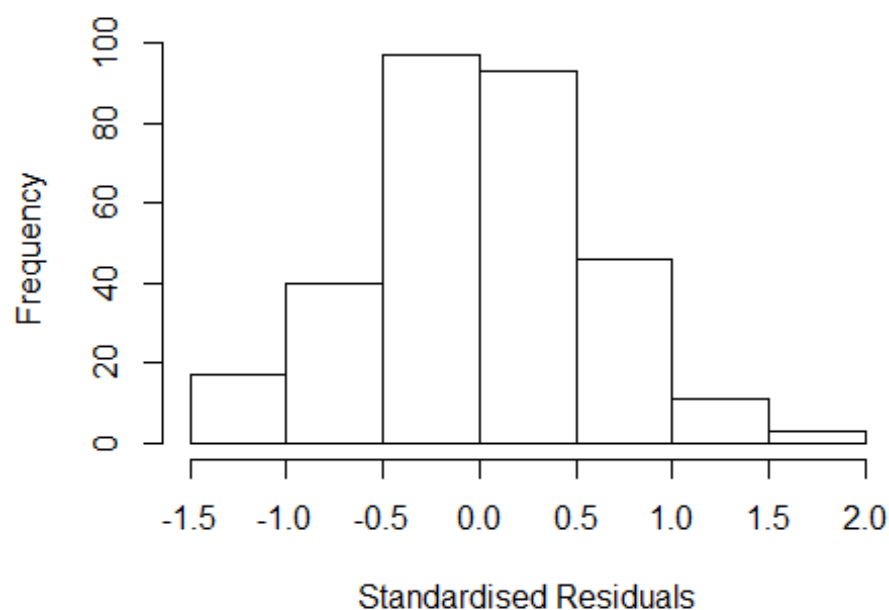
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31593 -0.35977 0.00538 0.37034 1.64238
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5309 0.3657 4.19 3.8e-05 ***
## FirstAuthorFemale1 -0.1263 0.1255 -1.01 0.315
## LastAuthorFemale1 -0.0922 0.1962 -0.47 0.639
## UniqueAuthors2 0.0499 0.0874 0.57 0.569
## UniqueAuthors3 0.1884 0.1137 1.66 0.099 .
## UniqueAuthors4 0.0288 0.1453 0.20 0.843
## UniqueAuthors5 0.2949 0.1913 1.54 0.124
## Year1997 -0.5606 0.3821 -1.47 0.143
## Year1998 0.2826 0.4149 0.68 0.496
## Year1999 -0.4032 0.3981 -1.01 0.312
```

```

## Year2000          -0.4033      0.3887    -1.04      0.300
## Year2001          -0.1562      0.3863    -0.40      0.686
## Year2002          -0.2391      0.4021    -0.59      0.553
## Year2003          -0.2152      0.3899    -0.55      0.581
## Year2004          -0.4144      0.3735    -1.11      0.268
## Year2005          -0.4664      0.4000    -1.17      0.245
## Year2006          -0.5318      0.3819    -1.39      0.165
## Year2007          -0.3360      0.3834    -0.88      0.382
## Year2008          -0.3138      0.4270    -0.73      0.463
## Year2009          -0.5985      0.3808    -1.57      0.117
## Year2010          -0.4776      0.3803    -1.26      0.210
## Year2011          -0.2198      0.3985    -0.55      0.582
## Year2012          -0.4571      0.3766    -1.21      0.226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.117, Adjusted R-squared:  0.0491
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 275 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.372  0.859  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      3.26e-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 1.643 1      1.282
## LastAuthorFemale  2.068 1      1.438
## Year              2.015 16      1.022

```

Residuals from first and last author



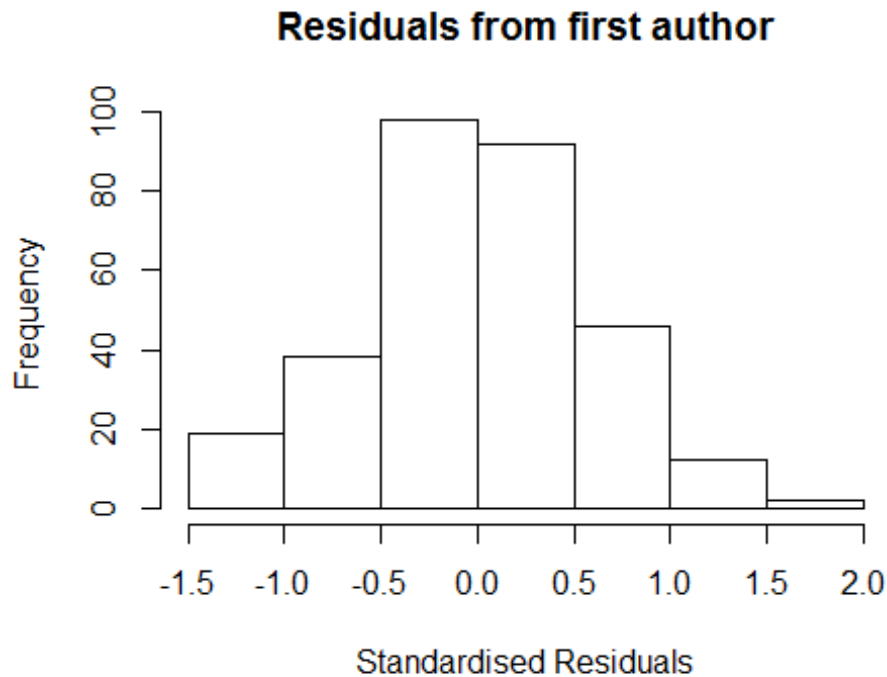
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40119 -0.36179 -0.00767  0.36271  1.76234
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5861     0.3414   4.65 5.2e-06 ***
## FirstAuthorFemale1 -0.0762     0.1247  -0.61   0.54
## LastAuthorFemale1 -0.0629     0.2032  -0.31   0.76
## Year1997          -0.5592     0.3553  -1.57   0.12
## Year1998           0.2647     0.3927   0.67   0.50
## Year1999          -0.3764     0.3727  -1.01   0.31
## Year2000          -0.4002     0.3688  -1.09   0.28
## Year2001          -0.1614     0.3648  -0.44   0.66
## Year2002          -0.2206     0.3826  -0.58   0.56
## Year2003          -0.2264     0.3692  -0.61   0.54
## Year2004          -0.3893     0.3494  -1.11   0.27
## Year2005          -0.4690     0.3723  -1.26   0.21
```



```

## Year2006          -0.5276      0.3584    -1.47      0.14
## Year2007          -0.3132      0.3584    -0.87      0.38
## Year2008          -0.2388      0.4023    -0.59      0.55
## Year2009          -0.5332      0.3545    -1.50      0.13
## Year2010          -0.4644      0.3590    -1.29      0.20
## Year2011          -0.1849      0.3752    -0.49      0.62
## Year2012          -0.4488      0.3522    -1.27      0.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.103, Adjusted R-squared:  0.0464
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 275 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.859  0.953  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      3.26e-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 1.217  1      1.103
## Year              1.217 16      1.006

```

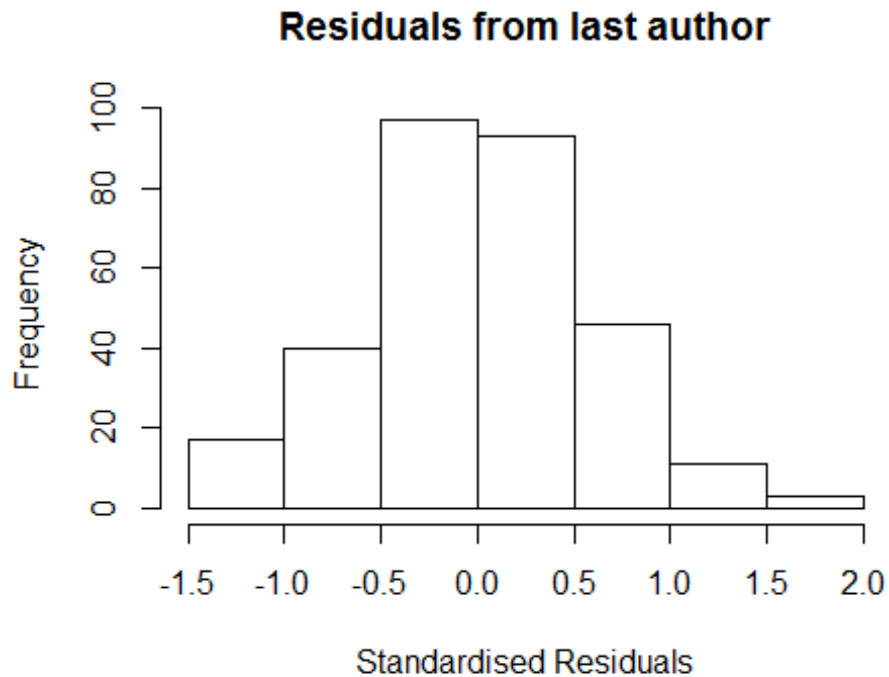


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40189 -0.36274 -0.00435 0.35941 1.76410
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5743 0.3353 4.69 4.1e-06 ***
## FirstAuthorFemale1 -0.0933 0.1093 -0.85 0.39
## Year1997 -0.5502 0.3501 -1.57 0.12
## Year1998 0.2766 0.3879 0.71 0.48
## Year1999 -0.3647 0.3670 -0.99 0.32
## Year2000 -0.3838 0.3603 -1.07 0.29
## Year2001 -0.1623 0.3661 -0.44 0.66
## Year2002 -0.2112 0.3786 -0.56 0.58
## Year2003 -0.2178 0.3635 -0.60 0.55
## Year2004 -0.3922 0.3496 -1.12 0.26
## Year2005 -0.4628 0.3687 -1.26 0.21
## Year2006 -0.5157 0.3495 -1.48 0.14
```

```

## Year2007          -0.3002      0.3493   -0.86      0.39
## Year2008          -0.2272      0.3942   -0.58      0.56
## Year2009          -0.5195      0.3472   -1.50      0.14
## Year2010          -0.4544      0.3522   -1.29      0.20
## Year2011          -0.1724      0.3703   -0.47      0.64
## Year2012          -0.4457      0.3508   -1.27      0.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0489
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 274 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.303  0.855  0.953  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      3.26e-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 1.608 1          1.268
## Year              1.608 16          1.015

```



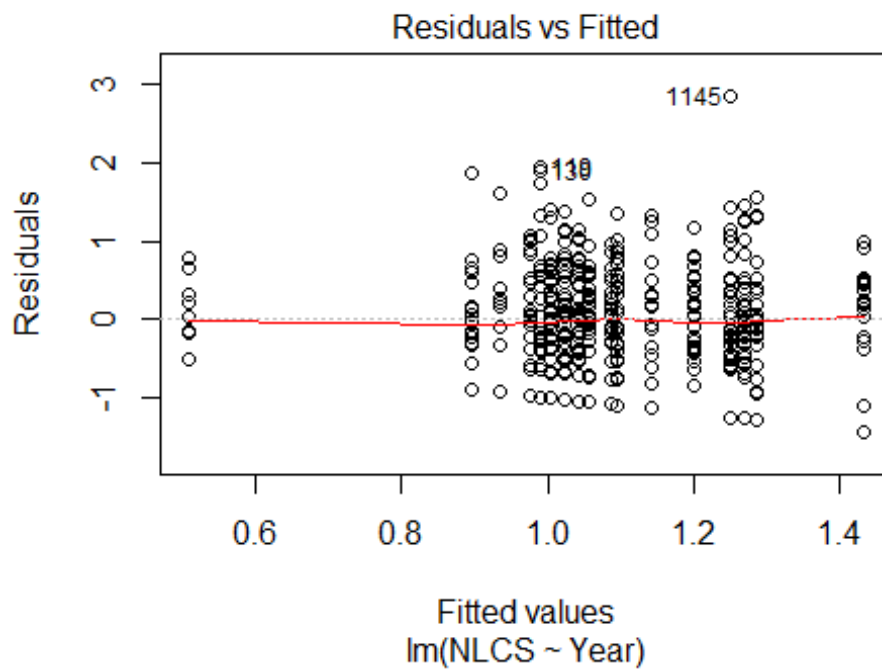
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.400 -0.361 -0.011 0.366 1.773
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5754 0.3408 4.62 5.7e-06 ***
## LastAuthorFemale1 -0.0969 0.1906 -0.51 0.61
## Year1997 -0.5574 0.3558 -1.57 0.12
## Year1998 0.2758 0.3923 0.70 0.48
## Year1999 -0.3657 0.3722 -0.98 0.33
## Year2000 -0.4088 0.3677 -1.11 0.27
## Year2001 -0.1537 0.3628 -0.42 0.67
## Year2002 -0.2183 0.3849 -0.57 0.57
## Year2003 -0.2193 0.3708 -0.59 0.55
## Year2004 -0.3708 0.3455 -1.07 0.28
## Year2005 -0.4564 0.3714 -1.23 0.22
## Year2006 -0.5280 0.3586 -1.47 0.14
```

```

## Year2007          -0.3075      0.3594   -0.86      0.39
## Year2008          -0.2424      0.4024   -0.60      0.55
## Year2009          -0.5306      0.3545   -1.50      0.14
## Year2010          -0.4639      0.3601   -1.29      0.20
## Year2011          -0.1753      0.3738   -0.47      0.64
## Year2012          -0.4431      0.3517   -1.26      0.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0494
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.296  0.857  0.953  0.901  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      3.26e-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: 307"
## [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
##   43   47   56   26   38   45   28   39   41   46   43   78   70   82   61
## 2011 2012
##   68   70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   15   33   13   23   22   18   21   32   30   19   48   43   38   32
## 2011 2012

```

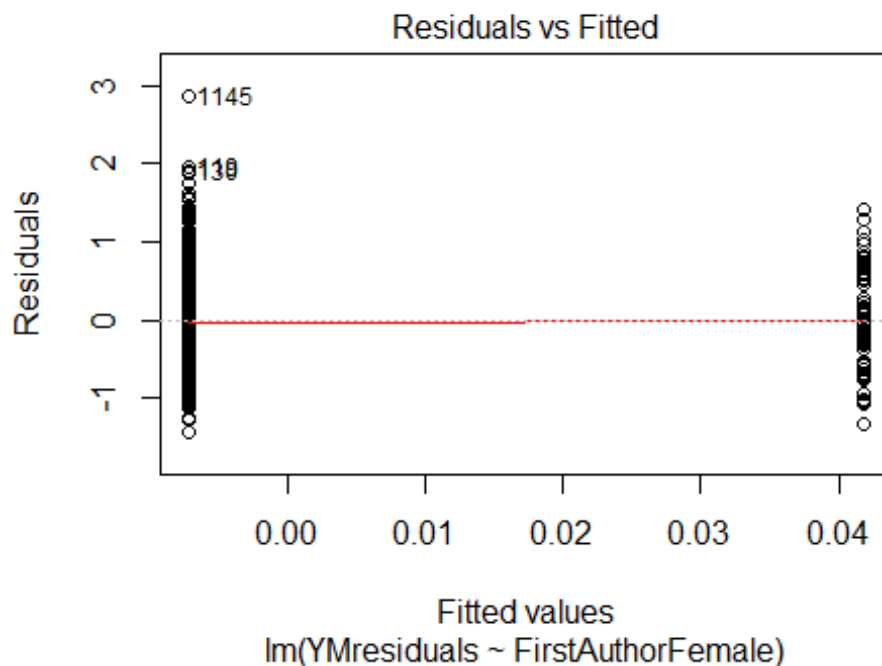
```
## 30 34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 13 27 11 22 14 11 18 26 27 14 40 37 31 26
## 2011 2012
## 21 22
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



```
##
## 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: 2.28942848510666"
## [1] "Male first author team size 2018 geometric mean: 2.52083204678585"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

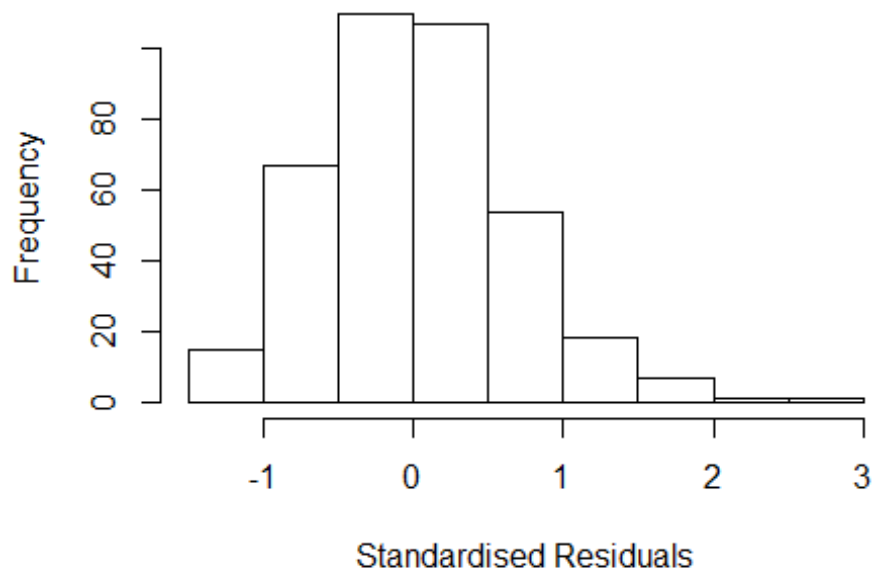
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 29, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.63901582154579"
## [1] "Male last author team size 2018 geometric mean: 2.4594185243188"

## 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 = 56, 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 1.542  1      1.242
## LastAuthorFemale  1.539  1      1.241
## UniqueAuthors    2.520  4      1.122
## Year              3.058 16      1.036
```

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
## 1145 84857370722 4.1 2012      1705      3      2.819
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28805 -0.40363 -0.00712  0.41703  2.81889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91108    0.16039     5.68 2.8e-08 ***
## FirstAuthorFemale1 -0.00360    0.09446    -0.04 0.96961
## LastAuthorFemale1  0.05349    0.09710     0.55 0.58203
## UniqueAuthors2     0.10394    0.09118     1.14 0.25511
## UniqueAuthors3     0.37239    0.10696     3.48 0.00056 ***
## UniqueAuthors4     0.54893    0.16976     3.23 0.00134 **
## UniqueAuthors5     0.47445    0.11577     4.10 5.2e-05 ***
## Year1997        -0.06425    0.29015    -0.22 0.82489
## Year1998        -0.15441    0.25728    -0.60 0.54879
## Year1999        -0.55060    0.21359    -2.58 0.01034 *
```

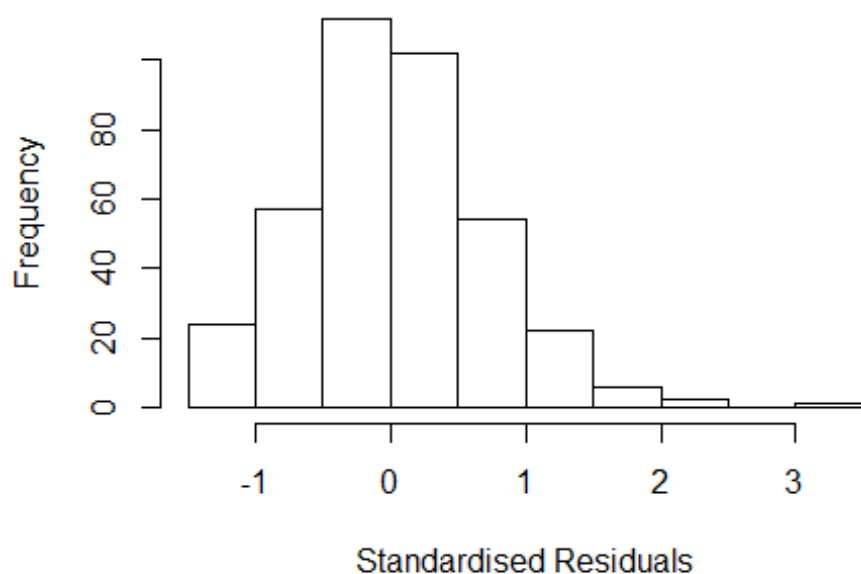


```

## Year2000      0.20415      0.22100      0.92  0.35623
## Year2001     -0.24823      0.22318     -1.11  0.26678
## Year2002      0.14104      0.36397      0.39  0.69862
## Year2003     -0.04649      0.20462     -0.23  0.82041
## Year2004     -0.07462      0.20793     -0.36  0.71991
## Year2005      0.07604      0.19294      0.39  0.69374
## Year2006      0.27304      0.21662      1.26  0.20834
## Year2007     -0.09452      0.17780     -0.53  0.59531
## Year2008     -0.15179      0.19116     -0.79  0.42772
## Year2009     -0.12406      0.19279     -0.64  0.52032
## Year2010     -0.08494      0.18170     -0.47  0.64045
## Year2011      0.09167      0.19718      0.46  0.64229
## Year2012     -0.00237      0.22845     -0.01  0.99174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0717
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 343 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0287 0.8630 0.9580 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          2.63e-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 1.303 1 1.141
## LastAuthorFemale 1.356 1 1.165
## Year 1.369 16 1.010

```

Residuals from first and last author

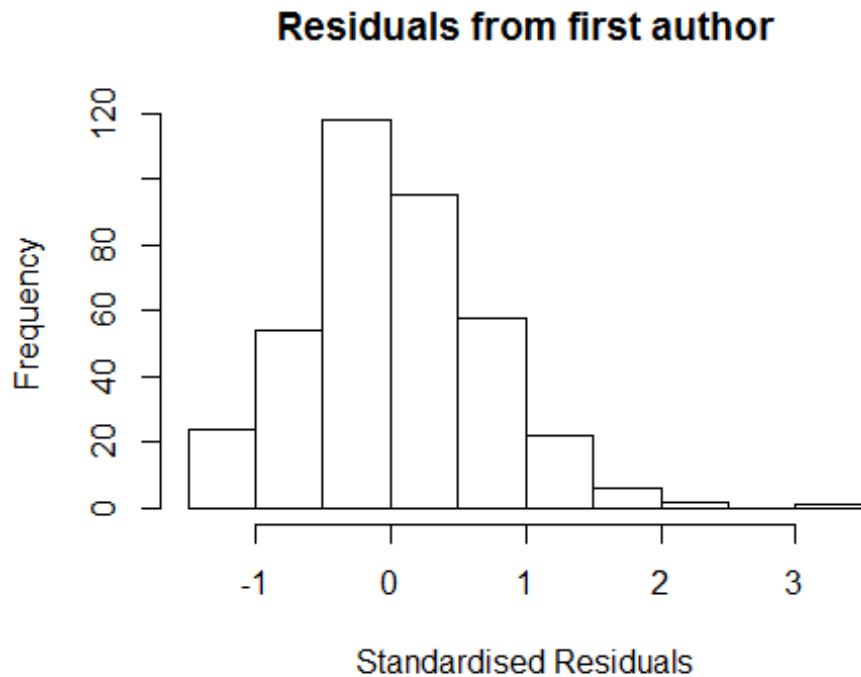


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 1145 84857370722 4.1 2012      1705      3      3.029
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4280 -0.4523 -0.0101  0.4591  3.0289
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.02521    0.15786     6.49  2.8e-10 ***
## FirstAuthorFemale1 -0.00698    0.09212    -0.08   0.9397
## LastAuthorFemale1  0.10137    0.09477     1.07   0.2855
## Year1997          -0.08486    0.28940    -0.29   0.7695
## Year1998          -0.19372    0.25936    -0.75   0.4556
## Year1999          -0.61518    0.20223    -3.04   0.0025 **
## Year2000           0.19840    0.23907     0.83   0.4072
## Year2001          -0.23619    0.22665    -1.04   0.2981
## Year2002           0.09214    0.38464     0.24   0.8108
## Year2003           0.00481    0.21209     0.02   0.9819
## Year2004           0.02292    0.20651     0.11   0.9117
## Year2005           0.14697    0.20092     0.73   0.4650
```

```

## Year2006          0.40284      0.22353      1.80      0.0724 .
## Year2007          -0.06553      0.18438     -0.36      0.7225
## Year2008          -0.05747      0.19395     -0.30      0.7672
## Year2009           0.00372      0.20563      0.02      0.9856
## Year2010           0.06455      0.18440      0.35      0.7265
## Year2011           0.16973      0.20287      0.84      0.4033
## Year2012           0.04588      0.23281      0.20      0.8439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.0641, Adjusted R-squared:  0.0175
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 354 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0054 0.8680 0.9540 0.9080 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      2.63e-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 1.15 1      1.073
## Year              1.15 16      1.004

```



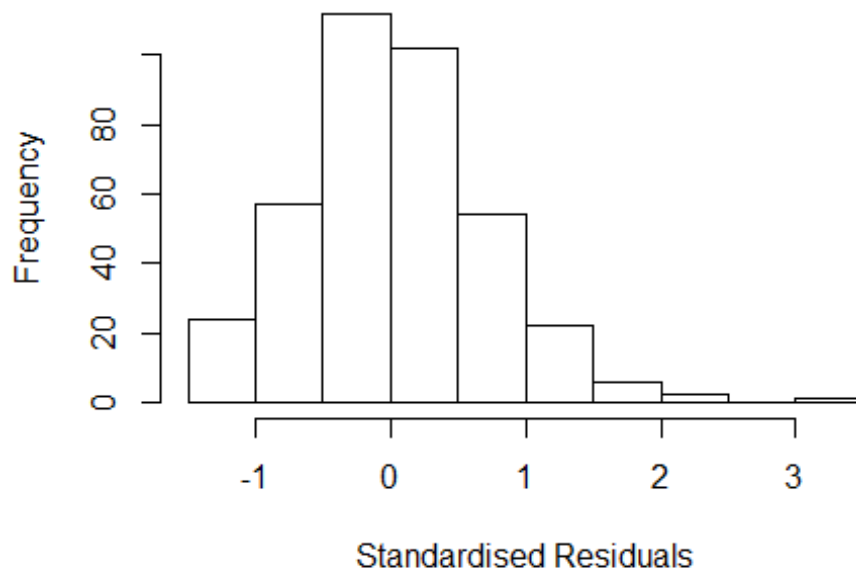
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 1145 84857370722 4.1 2012      1705      3      3.029
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4390 -0.4443 -0.0192  0.4423  3.0326
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0260     0.1585   6.47 3.1e-10 ***
## FirstAuthorFemale1  0.0263     0.0858   0.31  0.7590
## Year1997        -0.0855     0.2896  -0.30  0.7679
## Year1998        -0.1933     0.2599  -0.74  0.4576
## Year1999        -0.5815     0.1937  -3.00  0.0029 **
## Year2000         0.2152     0.2398   0.90  0.3700
## Year2001        -0.2253     0.2250  -1.00  0.3173
## Year2002         0.0866     0.3822   0.23  0.8210
## Year2003         0.0242     0.2090   0.12  0.9079
## Year2004         0.0399     0.2063   0.19  0.8469
## Year2005         0.1600     0.1987   0.81  0.4212
## Year2006         0.4130     0.2259   1.83  0.0683 .
```

```

## Year2007          -0.0628      0.1847   -0.34   0.7340
## Year2008          -0.0558      0.1954   -0.29   0.7754
## Year2009           0.0219      0.2041    0.11   0.9145
## Year2010           0.0733      0.1845    0.40   0.6913
## Year2011           0.1837      0.2021    0.91   0.3641
## Year2012           0.0413      0.2323    0.18   0.8589
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.0617, Adjusted R-squared:  0.0176
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0059 0.8710 0.9550 0.9090 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      2.63e-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 1.187 1          1.090
## Year              1.187 16          1.005

```

Residuals from last author



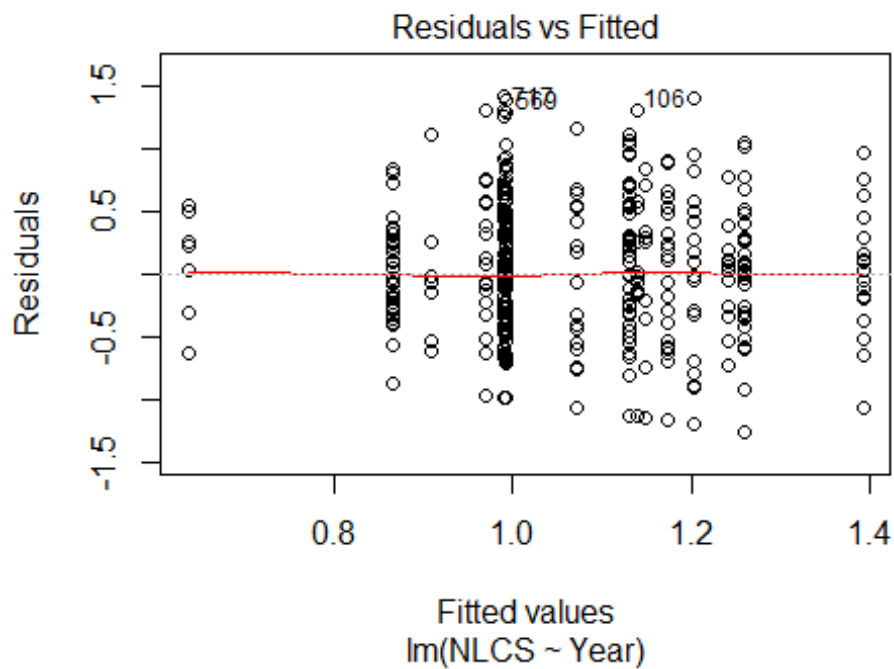
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 1145 84857370722 4.1 2012      1705      3      3.029
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42813 -0.45108 -0.00828  0.45926  3.03128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02419    0.15771     6.49  2.8e-10 ***
## LastAuthorFemale1 0.09884    0.08877     1.11  0.2663
## Year1997      -0.08404    0.28952    -0.29  0.7718
## Year1998      -0.19582    0.25938    -0.75  0.4508
## Year1999      -0.61570    0.20234    -3.04  0.0025 **
## Year2000       0.19834    0.23951     0.83  0.4082
## Year2001      -0.23635    0.22688    -1.04  0.2982
## Year2002       0.09167    0.38541     0.24  0.8121
## Year2003       0.00517    0.21246     0.02  0.9806
## Year2004       0.02380    0.20681     0.12  0.9084
## Year2005       0.14727    0.20033     0.74  0.4627
## Year2006       0.40394    0.22398     1.80  0.0721 .
```

```

## Year2007          -0.06540      0.18485    -0.35    0.7237
## Year2008          -0.05765      0.19427    -0.30    0.7668
## Year2009           0.00428      0.20611     0.02    0.9834
## Year2010           0.06500      0.18453     0.35    0.7249
## Year2011           0.16997      0.20282     0.84    0.4026
## Year2012           0.04453      0.23149     0.19    0.8476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.669
## Multiple R-squared:  0.0643, Adjusted R-squared:  0.0204
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0043 0.8660 0.9540 0.9080 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      2.63e-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: 380"
## [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
##   15   21   16    9   30   27   27   18   21   23   43   57   33   74   66
## 2011 2012
##   58   71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9    7   11    8   22   18   18   14   15   20   35   40   21   52   52
## 2011 2012

```

```
## 36 51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 6 10 7 20 17 16 13 12 19 32 36 19 46 44
## 2011 2012
## 29 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 = 19, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.8, df = 1, p-value = 0.4

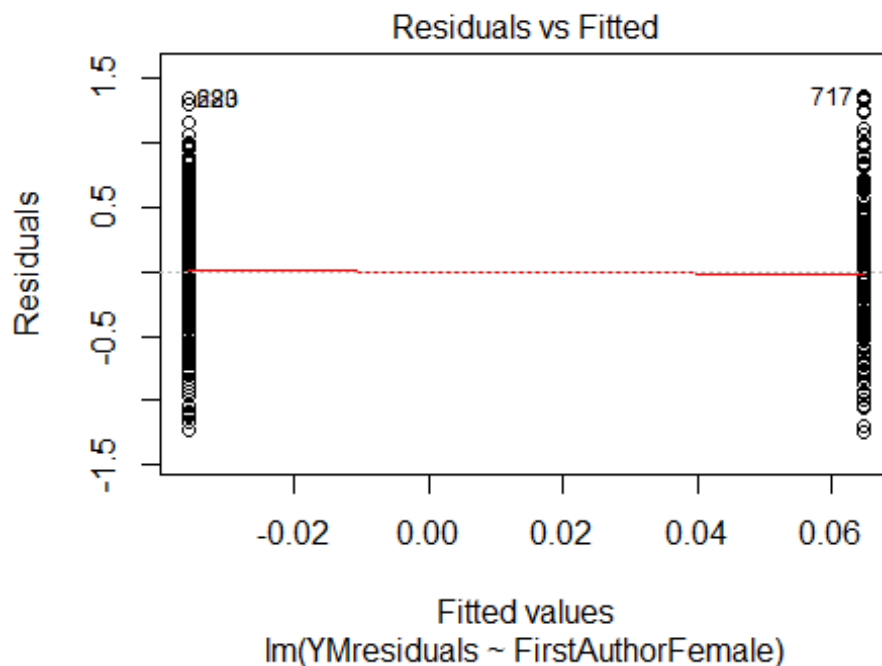
## [1] "Female first author team size 2018 geometric mean: 2.13644968540694"
## [1] "Male first author team size 2018 geometric mean: 2.6933955294532"

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



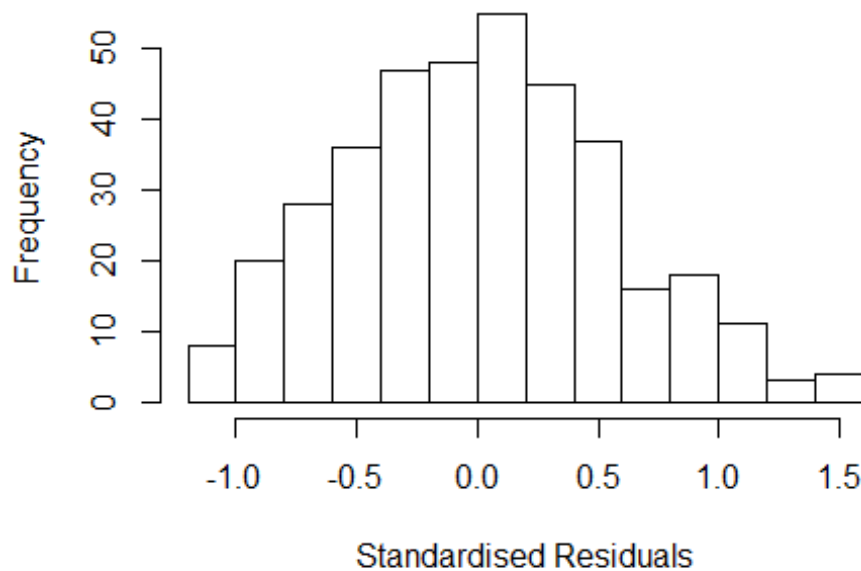
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 350, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.35302943932439"
## [1] "Male last author team size 2018 geometric mean: 2.56273012473719"

## 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 = 0.6
## 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 1.642  1      1.282
## LastAuthorFemale  1.589  1      1.260
## UniqueAuthors    1.880  4      1.082
## Year              2.460 16      1.029
```

Residuals from first and last author and team size



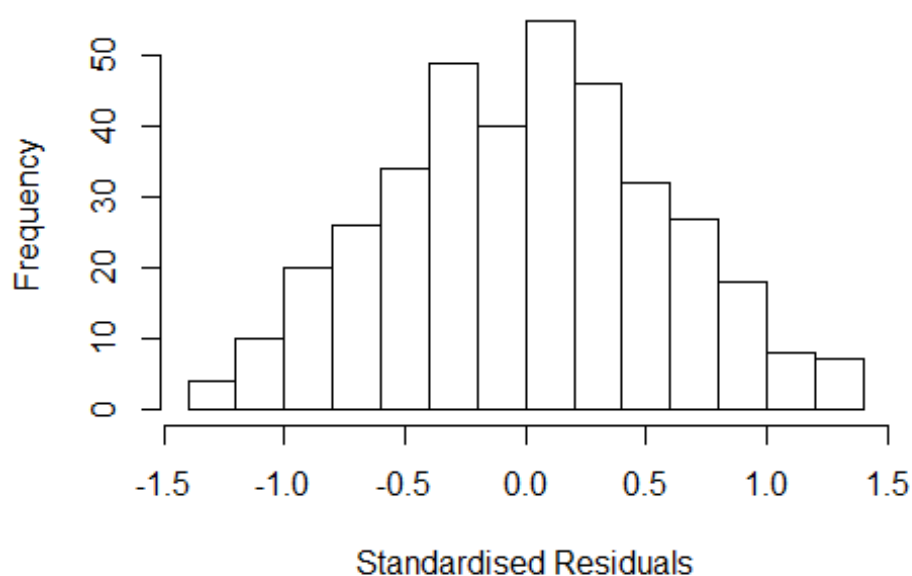
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11978 -0.39464 0.00066 0.36894 1.57278
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8313 0.3377 2.46 0.01429 *
## FirstAuthorFemale1 0.0888 0.0754 1.18 0.24016
## LastAuthorFemale1 0.0632 0.0754 0.84 0.40195
## UniqueAuthors2 0.2456 0.0777 3.16 0.00172 **
## UniqueAuthors3 0.4241 0.0847 5.01 8.8e-07 ***
## UniqueAuthors4 0.3851 0.1135 3.39 0.00077 ***
## UniqueAuthors5 0.5186 0.1770 2.93 0.00362 **
## Year1997 0.0787 0.3997 0.20 0.84402
## Year1998 0.0854 0.4020 0.21 0.83187
## Year1999 -0.3524 0.3762 -0.94 0.34950
```

```

## Year2000          0.0934      0.3581      0.26  0.79435
## Year2001          0.3620      0.3480      1.04  0.29902
## Year2002         -0.1810      0.3640     -0.50  0.61928
## Year2003          0.1073      0.3446      0.31  0.75584
## Year2004          0.0429      0.3703      0.12  0.90789
## Year2005         -0.1397      0.3587     -0.39  0.69711
## Year2006         -0.0908      0.3411     -0.27  0.79025
## Year2007         -0.2414      0.3352     -0.72  0.47195
## Year2008          0.0782      0.3566      0.22  0.82655
## Year2009          0.0436      0.3402      0.13  0.89816
## Year2010         -0.2135      0.3377     -0.63  0.52773
## Year2011          0.1544      0.3382      0.46  0.64837
## Year2012         -0.0734      0.3385     -0.22  0.82856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.159, Adjusted R-squared:  0.106
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 342 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.423  0.882  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      2.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.460 1      1.208
## LastAuthorFemale  1.425 1      1.194
## Year              1.355 16      1.010

```

Residuals from first and last author



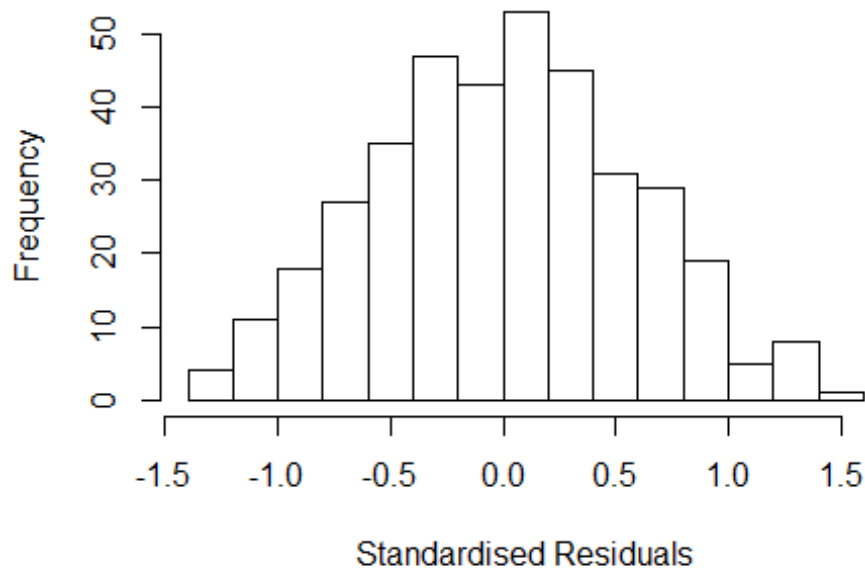
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2639 -0.3978 0.0208 0.3872 1.3839
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10283 0.28508 3.87 0.00013 ***
## FirstAuthorFemale1 0.08978 0.07721 1.16 0.24566
## LastAuthorFemale1 0.04898 0.07817 0.63 0.53133
## Year1997 -0.13625 0.35632 -0.38 0.70241
## Year1998 -0.08083 0.36684 -0.22 0.82574
## Year1999 -0.58565 0.33807 -1.73 0.08407 .
## Year2000 0.00934 0.33036 0.03 0.97747
## Year2001 0.23909 0.30776 0.78 0.43776
## Year2002 -0.18783 0.32705 -0.57 0.56612
## Year2003 0.07790 0.29840 0.26 0.79419
## Year2004 -0.03977 0.33928 -0.12 0.90675
## Year2005 -0.21187 0.32687 -0.65 0.51729
```

```

## Year2006      -0.17164      0.29528      -0.58      0.56141
## Year2007      -0.30045      0.29157      -1.03      0.30348
## Year2008        0.02235      0.31880        0.07      0.94415
## Year2009      -0.01107      0.29964      -0.04      0.97055
## Year2010      -0.23366      0.30147      -0.78      0.43882
## Year2011        0.11277      0.29885        0.38      0.70613
## Year2012      -0.13060      0.29975      -0.44      0.66332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0774, Adjusted R-squared:  0.0309
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 348 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.593  0.888  0.953  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      2.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.213 1      1.101
## Year              1.213 16      1.006

```

Residuals from first author



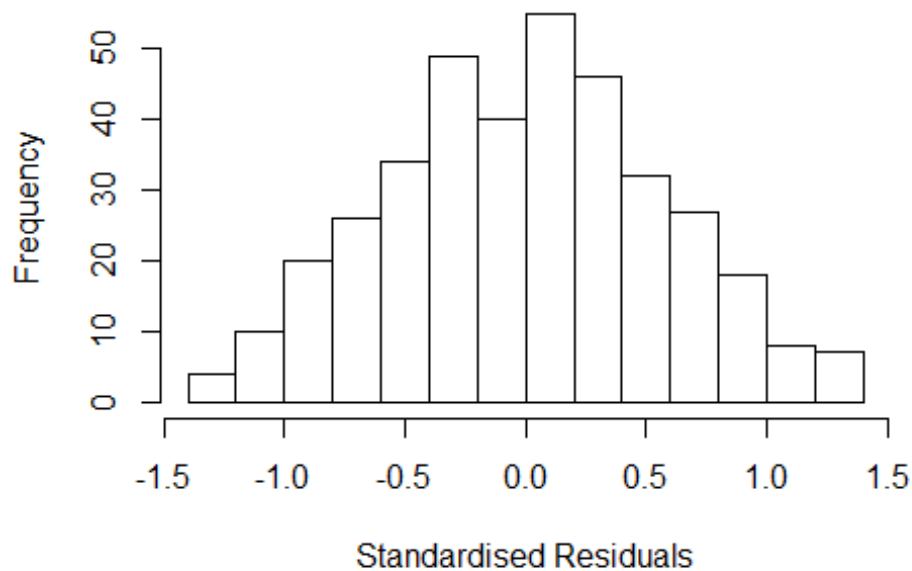
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24104 -0.40081 0.00753 0.39201 1.40111
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1038 0.2832 3.90 0.00012 ***
## FirstAuthorFemale1 0.1097 0.0698 1.57 0.11731
## Year1997 -0.1368 0.3547 -0.39 0.70001
## Year1998 -0.0759 0.3676 -0.21 0.83655
## Year1999 -0.5725 0.3361 -1.70 0.08937 .
## Year2000 0.0239 0.3267 0.07 0.94178
## Year2001 0.2433 0.3060 0.80 0.42704
## Year2002 -0.1798 0.3261 -0.55 0.58166
## Year2003 0.0871 0.2971 0.29 0.76969
## Year2004 -0.0247 0.3379 -0.07 0.94182
## Year2005 -0.1980 0.3217 -0.62 0.53856
## Year2006 -0.1680 0.2934 -0.57 0.56721
```

```

## Year2007          -0.2942      0.2894   -1.02   0.31002
## Year2008           0.0276      0.3180    0.09   0.93078
## Year2009           0.0033      0.2968    0.01   0.99113
## Year2010          -0.2245      0.2999   -0.75   0.45457
## Year2011           0.1235      0.2963    0.42   0.67698
## Year2012          -0.1263      0.2979   -0.42   0.67183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0765, Adjusted R-squared:  0.0326
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.588  0.888  0.956  0.921  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.194 1      1.093
## Year              1.194 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
## -1.24478 -0.38182 0.00515 0.40315 1.42710
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12575 0.27420 4.11 5e-05 ***
## LastAuthorFemale1 0.08672 0.07108 1.22 0.223
## Year1997 -0.15847 0.34763 -0.46 0.649
## Year1998 -0.08240 0.36450 -0.23 0.821
## Year1999 -0.59932 0.32395 -1.85 0.065 .
## Year2000 0.00323 0.32350 0.01 0.992
## Year2001 0.23705 0.30093 0.79 0.431
## Year2002 -0.17517 0.32153 -0.54 0.586
## Year2003 0.07968 0.29068 0.27 0.784
## Year2004 -0.04236 0.33066 -0.13 0.898
## Year2005 -0.21753 0.31747 -0.69 0.494
## Year2006 -0.17635 0.28673 -0.62 0.539
```

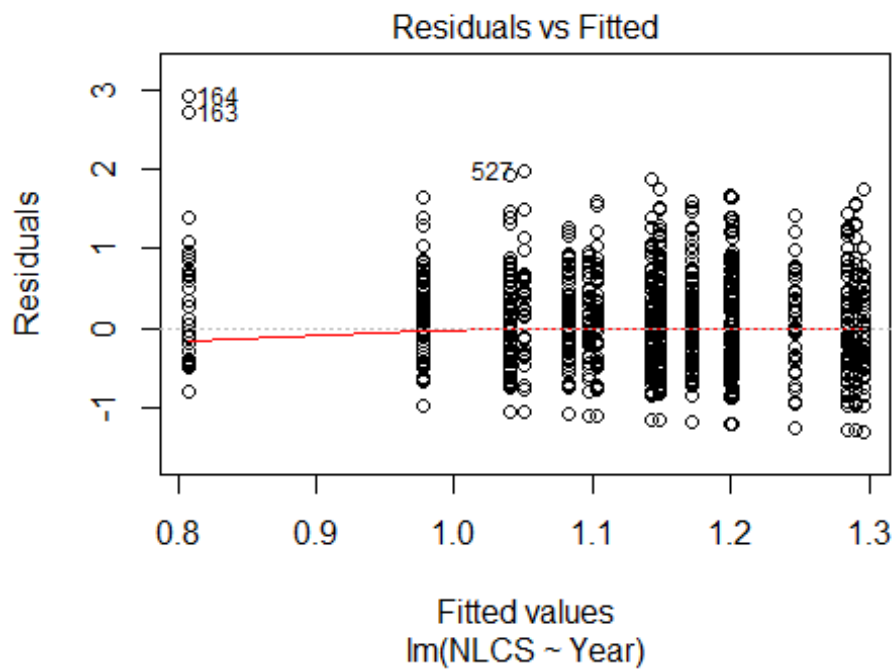


```

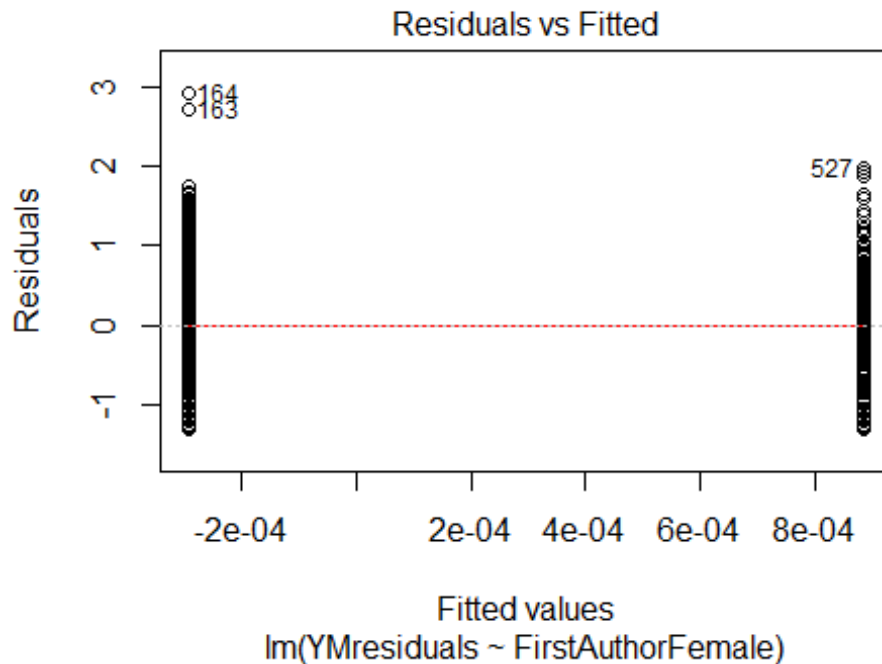
## Year2007          -0.30017      0.28302    -1.06      0.290
## Year2008           0.00785      0.30856      0.03      0.980
## Year2009          -0.01048      0.29170     -0.04      0.971
## Year2010          -0.24957      0.29181     -0.86      0.393
## Year2011           0.11903      0.29078      0.41      0.683
## Year2012          -0.13340      0.29145     -0.46      0.647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.0735, Adjusted R-squared:  0.0295
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 349 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.578  0.892  0.957   0.921  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 376"
## [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
##   66   84   66   77   72   76   73   70   65   96  123  136  127  160  144
## 2011 2012
##  125  149
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   45   40   51   48   38   51   42   49   63   81   90   97  101  100
## 2011 2012

```

```
##      82  103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32   41   32   44   44   36   41   36   42   56   68   75   86   88   81
## 2011 2012
##   63   76
## [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
```

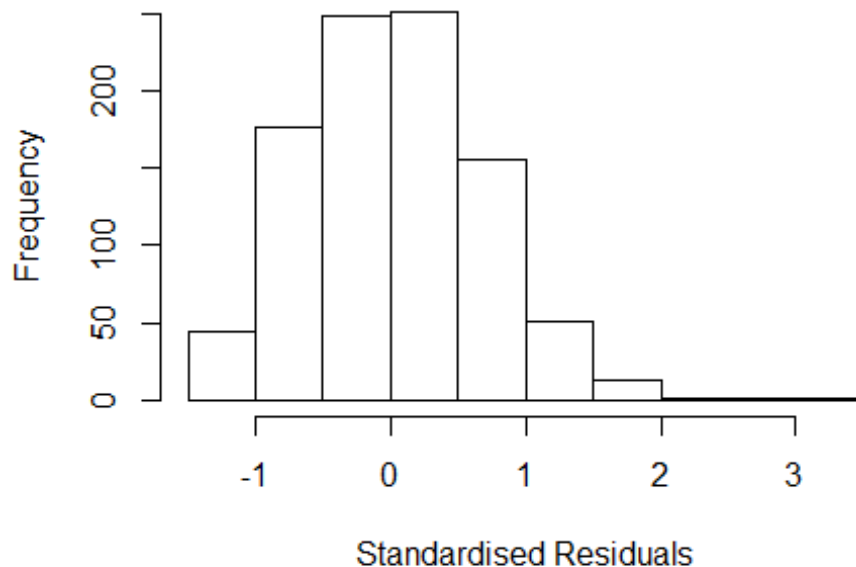


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.61, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.51747162269296"
## [1] "Male first author team size 2018 geometric mean: 2.56755004817431"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 720, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.31435793279335"
## [1] "Male last author team size 2018 geometric mean: 2.62748398801783"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 550, 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 1.641 1      1.281
## LastAuthorFemale  1.614 1      1.270
## UniqueAuthors     1.304 4      1.034
## Year               1.500 16     1.013
```

Residuals from first and last author and team size



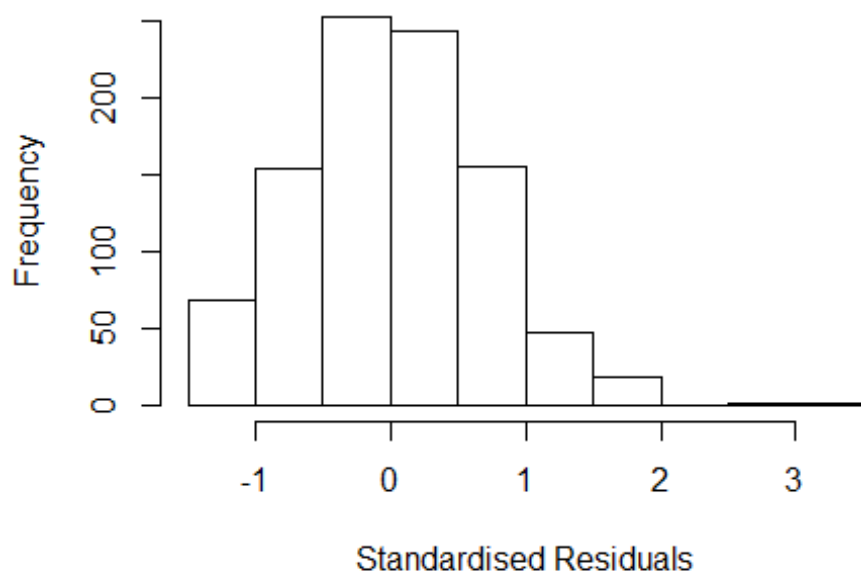
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 163 0031435009 3.522 1997    1710      3      2.595
## 164 0031482189 3.717 1997    1705      4      3.149
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.46701 -0.44846  0.00812  0.47624  3.14870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96117    0.15183     6.33 3.8e-10 ***
## FirstAuthorFemale1 0.00931    0.06103     0.15  0.879
## LastAuthorFemale1 -0.01661    0.06291    -0.26  0.792
## UniqueAuthors2    0.24184    0.05318     4.55 6.2e-06 ***
## UniqueAuthors3    0.35846    0.06102     5.87 5.9e-09 ***
## UniqueAuthors4    0.41403    0.10455     3.96 8.1e-05 ***
## UniqueAuthors5    0.31379    0.14315     2.19  0.029 *
## Year1997        -0.39286    0.19844    -1.98  0.048 *
## Year1998         0.13866    0.19450     0.71  0.476
```

```

## Year1999      -0.00895    0.20061   -0.04    0.964
## Year2000      0.18380    0.18423    1.00    0.319
## Year2001     -0.06209    0.20677   -0.30    0.764
## Year2002      0.14738    0.19592    0.75    0.452
## Year2003      0.11544    0.18377    0.63    0.530
## Year2004     -0.05228    0.17481   -0.30    0.765
## Year2005      0.09945    0.17872    0.56    0.578
## Year2006     -0.13556    0.17442   -0.78    0.437
## Year2007     -0.02633    0.17062   -0.15    0.877
## Year2008     -0.00674    0.16913   -0.04    0.968
## Year2009     -0.22295    0.16736   -1.33    0.183
## Year2010     -0.05088    0.17186   -0.30    0.767
## Year2011     -0.10945    0.17113   -0.64    0.523
## Year2012      0.00703    0.17943    0.04    0.969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.665
## Multiple R-squared:  0.0852, Adjusted R-squared:  0.0632
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 68 is an outlier with |weight| = 0 ( < 0.00011);
## 70 weights are ~= 1. The remaining 870 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0939 0.8760 0.9480 0.9120 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 1.561 1 1.250
## LastAuthorFemale 1.584 1 1.259
## Year 1.208 16 1.006

```

Residuals from first and last author



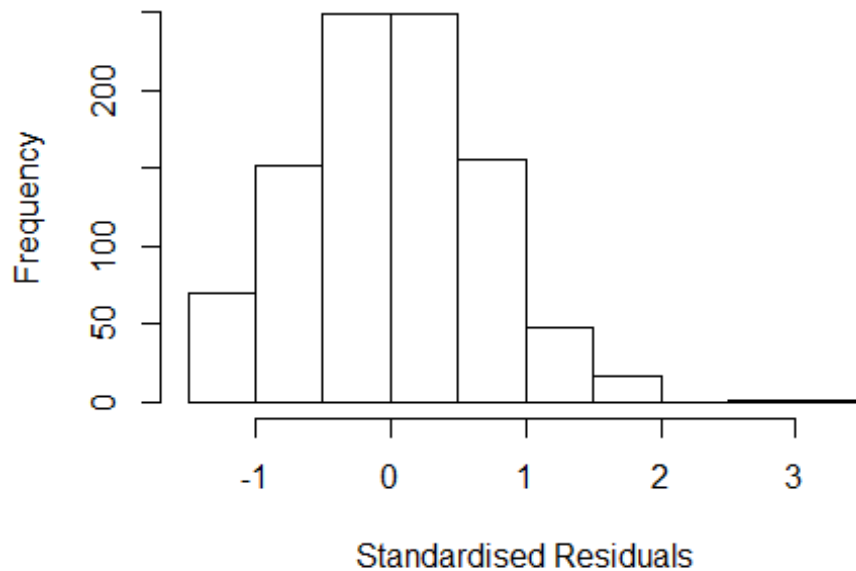
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 163 0031435009 3.522 1997    1710      3      2.853
## 164 0031482189 3.717 1997    1705      4      3.048
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.28338 -0.46932 -0.00548  0.46370  3.04828
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0810     0.1557   6.94 7.2e-12 ***
## FirstAuthorFemale1  0.0250     0.0616   0.41  0.684
## LastAuthorFemale1 -0.0272     0.0645  -0.42  0.674
## Year1997          -0.4123     0.1995  -2.07  0.039 *
## Year1998           0.1504     0.1970   0.76  0.446
## Year1999          -0.0133     0.2003  -0.07  0.947
## Year2000           0.1836     0.1925   0.95  0.340
## Year2001          -0.0296     0.2112  -0.14  0.889
## Year2002           0.1696     0.1993   0.85  0.395
## Year2003           0.2045     0.1887   1.08  0.279
## Year2004           0.0120     0.1795   0.07  0.947
```

```

## Year2005          0.1768      0.1826      0.97      0.333
## Year2006         -0.0610      0.1759     -0.35      0.729
## Year2007          0.0425      0.1740      0.24      0.807
## Year2008          0.0505      0.1723      0.29      0.770
## Year2009         -0.1437      0.1697     -0.85      0.397
## Year2010          0.0513      0.1733      0.30      0.767
## Year2011         -0.0334      0.1764     -0.19      0.850
## Year2012          0.1294      0.1797      0.72      0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0385, Adjusted R-squared:  0.0197
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 865 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0108 0.8830 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      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 1.088 1      1.043
## Year              1.088 16      1.003

```

Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 163 0031435009 3.522 1997    1710     3     2.853
## 164 0031482189 3.717 1997    1705     4     3.048
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29459 -0.47363  0.00012  0.46091  3.05137
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08154    0.15538   6.96 6.4e-12 ***
## FirstAuthorFemale1 0.01129    0.05150   0.22  0.827
## Year1997      -0.41591    0.19906  -2.09  0.037 *
## Year1998       0.14768    0.19660   0.75  0.453
## Year1999      -0.01773    0.20050  -0.09  0.930
## Year2000       0.18276    0.19242   0.95  0.342
## Year2001      -0.03260    0.21059  -0.15  0.877
## Year2002       0.16365    0.19815   0.83  0.409
## Year2003       0.20176    0.18824   1.07  0.284
## Year2004       0.00538    0.17782   0.03  0.976
## Year2005       0.17439    0.18187   0.96  0.338
```

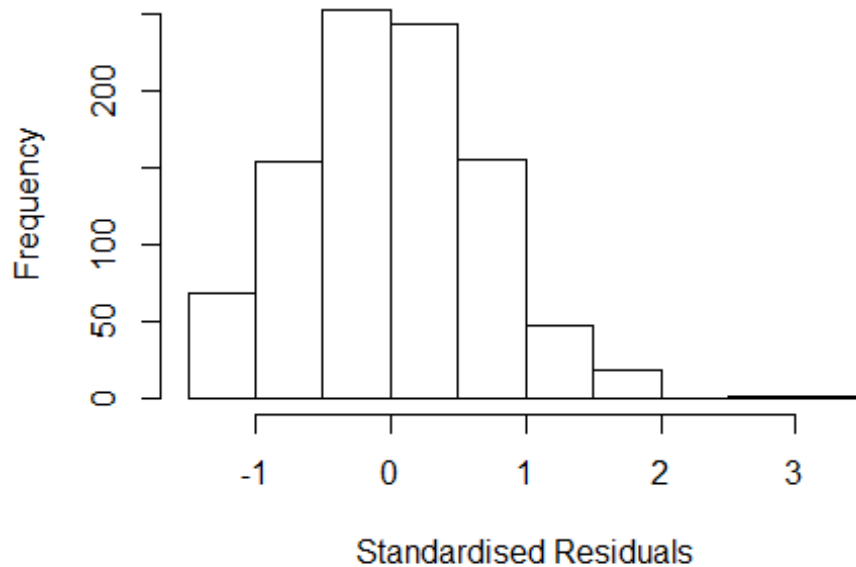


```

## Year2006      -0.06248    0.17583   -0.36    0.722
## Year2007      0.04157    0.17385    0.24    0.811
## Year2008      0.04804    0.17162    0.28    0.780
## Year2009     -0.15031    0.16854   -0.89    0.373
## Year2010      0.04755    0.17250    0.28    0.783
## Year2011     -0.03791    0.17535   -0.22    0.829
## Year2012      0.12634    0.17922    0.70    0.481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.69
## Multiple R-squared:  0.0382, Adjusted R-squared:  0.0205
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 866 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.012  0.884  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      1.06e-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 1.107 1      1.052
## Year      1.107 16      1.003

```

Residuals from last author



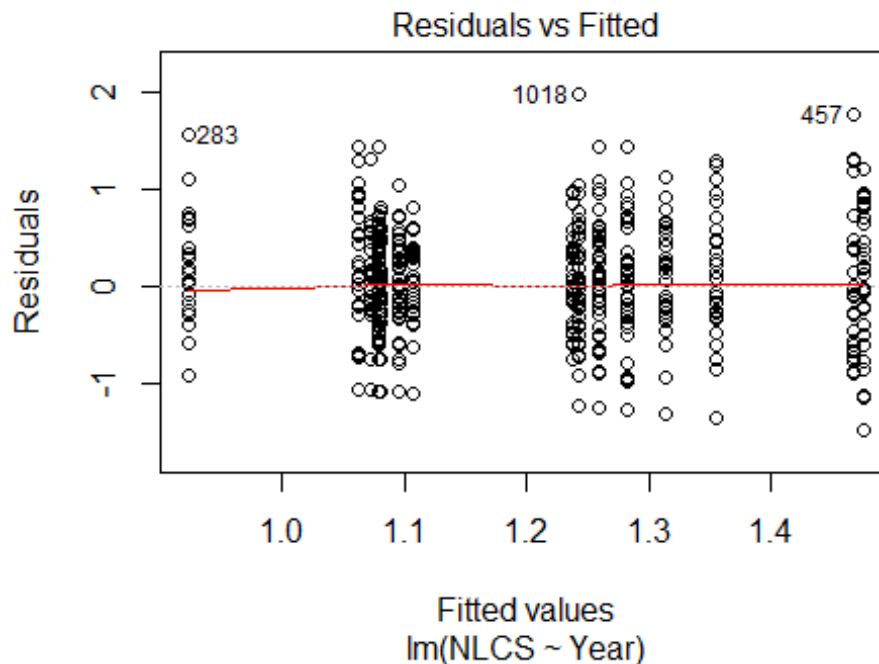
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 163 0031435009 3.522 1997    1710     3    2.853
## 164 0031482189 3.717 1997    1705     4    3.048
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.27724 -0.47305  0.00427  0.46458  3.04791
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0832     0.1560   6.94 7.2e-12 ***
## LastAuthorFemale1 -0.0137     0.0539  -0.25  0.799
## Year1997         -0.4141     0.1997  -2.07  0.038 *
## Year1998          0.1488     0.1973   0.75  0.451
## Year1999         -0.0134     0.2005  -0.07  0.947
## Year2000          0.1858     0.1922   0.97  0.334
## Year2001         -0.0288     0.2115  -0.14  0.892
## Year2002          0.1673     0.1992   0.84  0.401
## Year2003          0.2078     0.1883   1.10  0.270
## Year2004          0.0117     0.1797   0.06  0.948
## Year2005          0.1772     0.1828   0.97  0.332
```

```

## Year2006          -0.0596      0.1759    -0.34      0.735
## Year2007           0.0471      0.1735      0.27      0.786
## Year2008           0.0515      0.1724      0.30      0.765
## Year2009          -0.1434      0.1699    -0.84      0.399
## Year2010           0.0531      0.1734      0.31      0.760
## Year2011          -0.0322      0.1764    -0.18      0.855
## Year2012           0.1327      0.1795      0.74      0.460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.0206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 865 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0104 0.8820 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      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: 941"
## [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
##   56   74   77   62   59   61   57   50   43   48   57   81   75   79   69
## 2011 2012
##   65   88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   33   35   26   21   25   26   23   24   31   33   37   41   39   41

```

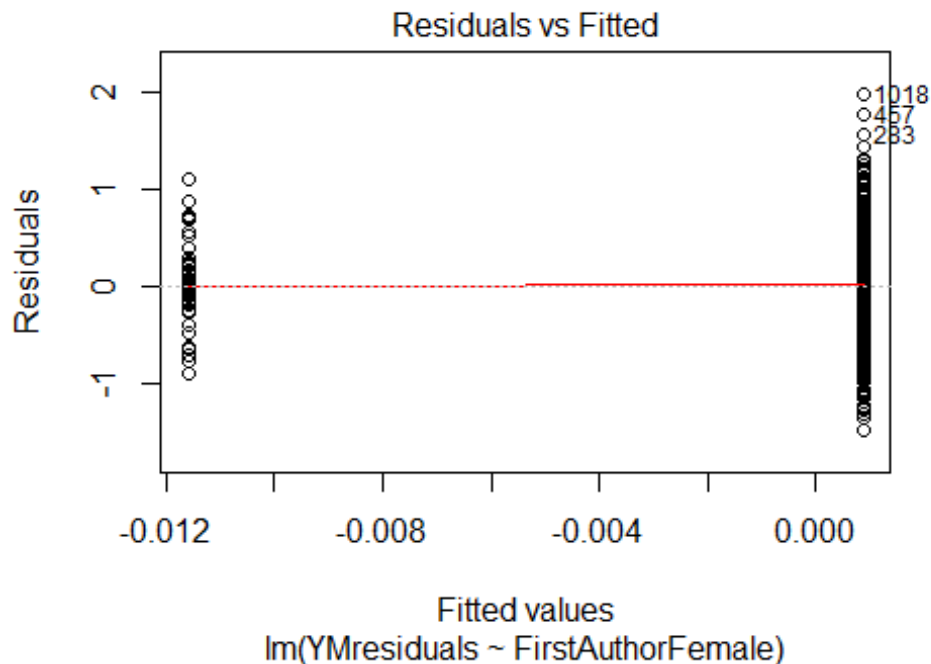
```
## 2011 2012
## 29 56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 23 22 24 15 16 13 20 18 18 28 27 26 33 33 30
## 2011 2012
## 26 41
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## 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 = 3.5, df = 1, p-value = 0.06
## [1] "Female first author team size 2018 geometric mean: 4.55901411390956"
## [1] "Male first author team size 2018 geometric mean: 2.37449846123539"
## 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 = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.89897948556636"
## [1] "Male last author team size 2018 geometric mean: 2.35388840791427"

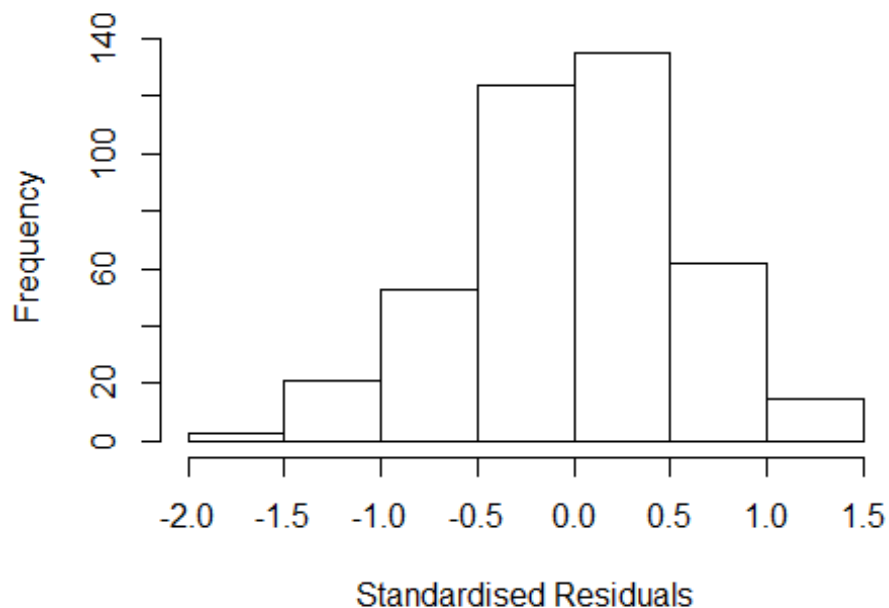
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.06
## 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	1.383	1	1.176
LastAuthorFemale	1.386	1	1.177
UniqueAuthors	2.813	4	1.138
Year	3.178	16	1.037

Residuals from first and last author and team size



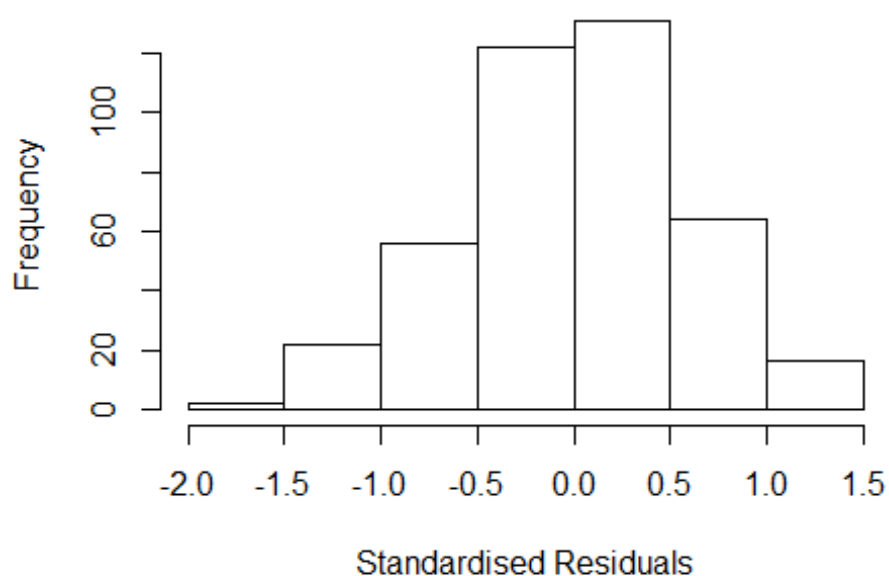
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7241 -0.3498 0.0129 0.3682 1.4622
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45905 0.26043 5.60 4e-08 ***
## FirstAuthorFemale1 0.00367 0.10529 0.03 0.972
## LastAuthorFemale1 -0.03240 0.14950 -0.22 0.829
## UniqueAuthors2 -0.03799 0.07189 -0.53 0.598
## UniqueAuthors3 0.01645 0.08474 0.19 0.846
## UniqueAuthors4 0.29302 0.14869 1.97 0.049 *
## UniqueAuthors5 0.09509 0.18092 0.53 0.599
## Year1997 -0.52928 0.30358 -1.74 0.082 .
## Year1998 0.26505 0.29623 0.89 0.371
## Year1999 -0.61621 0.30521 -2.02 0.044 *
```

```

## Year2000      -0.40522    0.29744   -1.36    0.174
## Year2001      -0.08636    0.34011   -0.25    0.800
## Year2002      -0.43400    0.30930   -1.40    0.161
## Year2003      -0.16793    0.27652   -0.61    0.544
## Year2004      -0.15933    0.31455   -0.51    0.613
## Year2005      -0.39016    0.27224   -1.43    0.153
## Year2006      -0.28086    0.26747   -1.05    0.294
## Year2007      -0.20629    0.27218   -0.76    0.449
## Year2008      -0.21950    0.27860   -0.79    0.431
## Year2009      -0.38062    0.27965   -1.36    0.174
## Year2010      -0.35560    0.27729   -1.28    0.200
## Year2011      -0.19316    0.29318   -0.66    0.510
## Year2012      -0.30528    0.28272   -1.08    0.281
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.109, Adjusted R-squared:  0.0591
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 372 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.328  0.864  0.953  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      2.42e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.282 1      1.132
## LastAuthorFemale  1.259 1      1.122
## Year              1.368 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
## -1.7071 -0.3529 0.0178 0.3573 1.4913
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4372 0.2462 5.84 1.1e-08 ***
## FirstAuthorFemale1 0.0251 0.1009 0.25 0.804
## LastAuthorFemale1 -0.0574 0.1448 -0.40 0.692
## Year1997 -0.5239 0.2907 -1.80 0.072 .
## Year1998 0.2699 0.2864 0.94 0.347
## Year1999 -0.6054 0.2951 -2.05 0.041 *
## Year2000 -0.3795 0.2825 -1.34 0.180
## Year2001 -0.0813 0.3292 -0.25 0.805
## Year2002 -0.3971 0.2824 -1.41 0.160
## Year2003 -0.1253 0.2661 -0.47 0.638
## Year2004 -0.1064 0.3037 -0.35 0.726
## Year2005 -0.3779 0.2596 -1.46 0.146
```

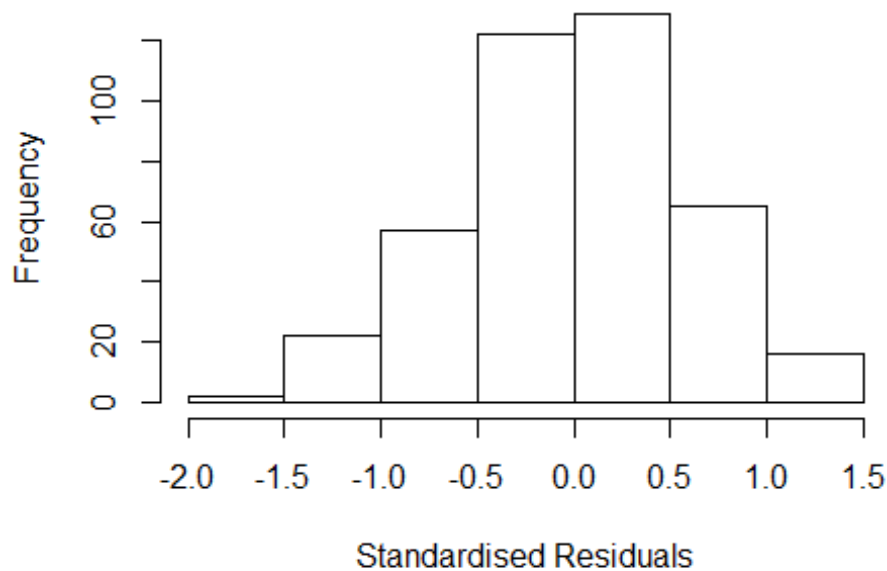


```

## Year2006          -0.2715      0.2547    -1.07     0.287
## Year2007          -0.1520      0.2574    -0.59     0.555
## Year2008          -0.1921      0.2642    -0.73     0.468
## Year2009          -0.2987      0.2598    -1.15     0.251
## Year2010          -0.3357      0.2631    -1.28     0.203
## Year2011          -0.1437      0.2786    -0.52     0.606
## Year2012          -0.2455      0.2700    -0.91     0.364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.0925, Adjusted R-squared:  0.0511
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 376 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.351  0.871  0.952  0.903  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      2.42e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.173 1          1.083
## Year              1.173 16          1.005

```

Residuals from first author



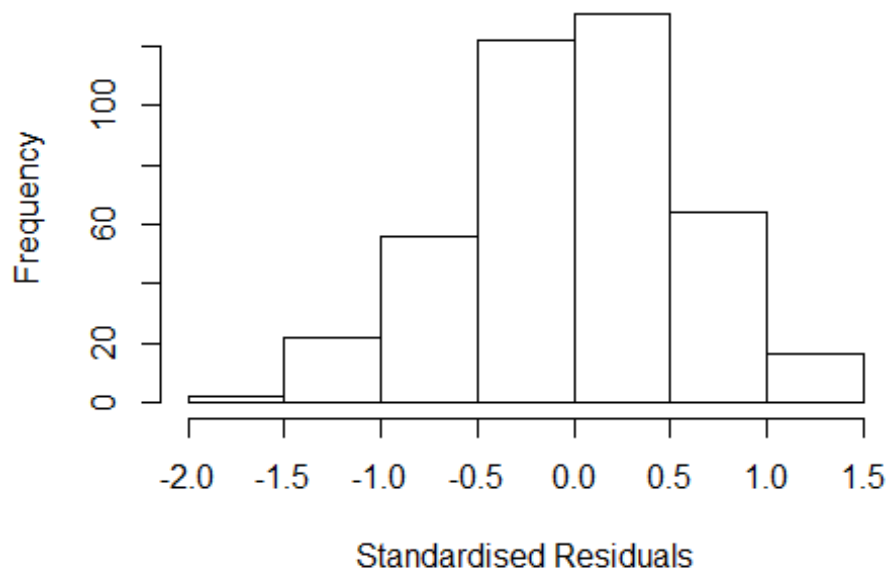
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7084 -0.3479 0.0174 0.3589 1.4929
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4355 0.2483 5.78 1.5e-08 ***
## FirstAuthorFemale1 0.0130 0.0969 0.13 0.894
## Year1997 -0.5225 0.2915 -1.79 0.074 .
## Year1998 0.2728 0.2878 0.95 0.344
## Year1999 -0.6054 0.2979 -2.03 0.043 *
## Year2000 -0.3759 0.2841 -1.32 0.187
## Year2001 -0.0894 0.3329 -0.27 0.788
## Year2002 -0.3977 0.2837 -1.40 0.162
## Year2003 -0.1287 0.2680 -0.48 0.631
## Year2004 -0.1179 0.3010 -0.39 0.696
## Year2005 -0.3781 0.2614 -1.45 0.149
## Year2006 -0.2707 0.2565 -1.06 0.292
```

```

## Year2007          -0.1549      0.2589   -0.60    0.550
## Year2008          -0.1940      0.2660   -0.73    0.466
## Year2009          -0.2978      0.2618   -1.14    0.256
## Year2010          -0.3404      0.2651   -1.28    0.200
## Year2011          -0.1471      0.2797   -0.53    0.599
## Year2012          -0.2454      0.2719   -0.90    0.367
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.092, Adjusted R-squared:  0.0529
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 377 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.348  0.874  0.952  0.903  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      2.42e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.155 1          1.075
## Year            1.155 16          1.005

```

Residuals from last author



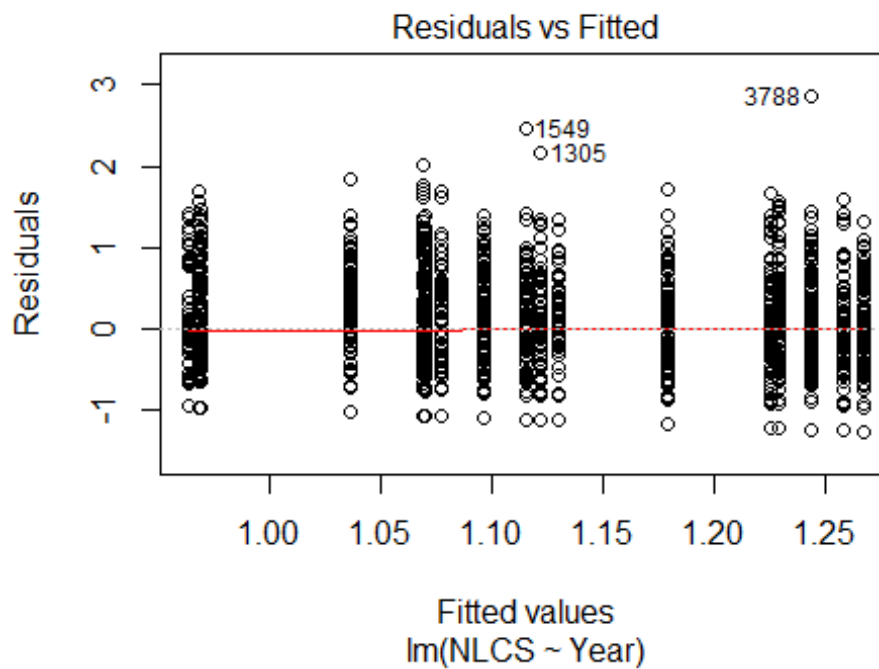
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7090 -0.3499 0.0161 0.3607 1.4891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4400 0.2469 5.83 1.1e-08 ***
## LastAuthorFemale1 -0.0516 0.1393 -0.37 0.712
## Year1997 -0.5240 0.2914 -1.80 0.073 .
## Year1998 0.2690 0.2873 0.94 0.350
## Year1999 -0.6065 0.2966 -2.04 0.042 *
## Year2000 -0.3784 0.2837 -1.33 0.183
## Year2001 -0.0853 0.3296 -0.26 0.796
## Year2002 -0.3970 0.2834 -1.40 0.162
## Year2003 -0.1278 0.2666 -0.48 0.632
## Year2004 -0.1105 0.3033 -0.36 0.716
## Year2005 -0.3802 0.2603 -1.46 0.145
## Year2006 -0.2715 0.2560 -1.06 0.289
```

```

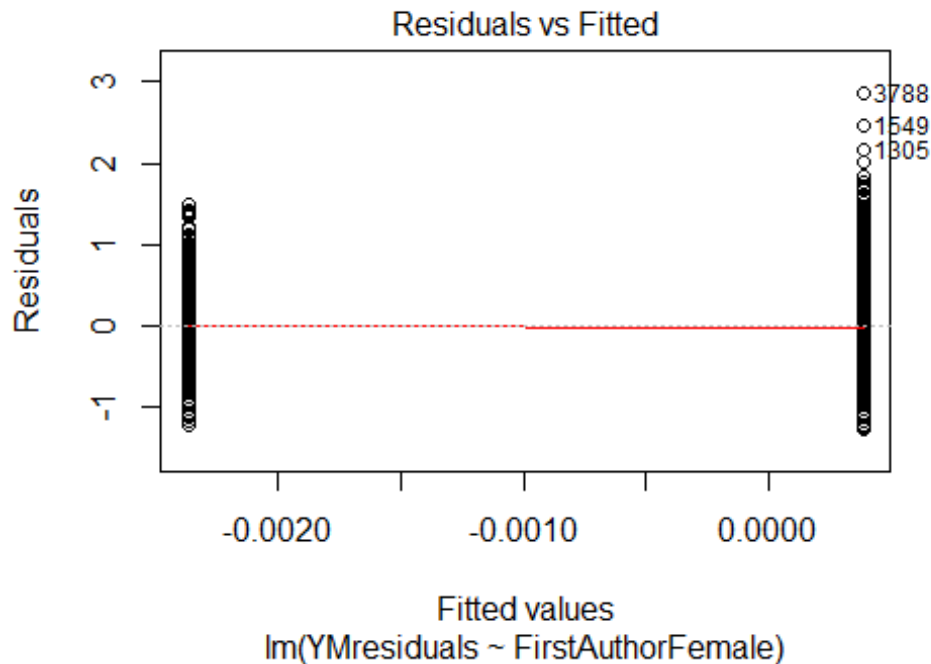
## Year2007          -0.1533      0.2586   -0.59    0.554
## Year2008          -0.1911      0.2658   -0.72    0.473
## Year2009          -0.3007      0.2609   -1.15    0.250
## Year2010          -0.3386      0.2634   -1.29    0.199
## Year2011          -0.1469      0.2789   -0.53    0.599
## Year2012          -0.2461      0.2713   -0.91    0.365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.0925, Adjusted R-squared:  0.0535
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 375 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.348  0.871  0.952  0.903  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      2.42e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 413"
## [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
## 149 159 154 139 147 154 119 115 132 168 171 219 219 256 204
## 2011 2012
## 222 258
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 76 82 62 74 74 77 60 80 111 94 136 126 130 107
## 2011 2012

```

```
## 113 140
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 64 67 62 50 67 62 60 52 64 99 75 116 108 110 87
## 2011 2012
## 90 105
## [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
```

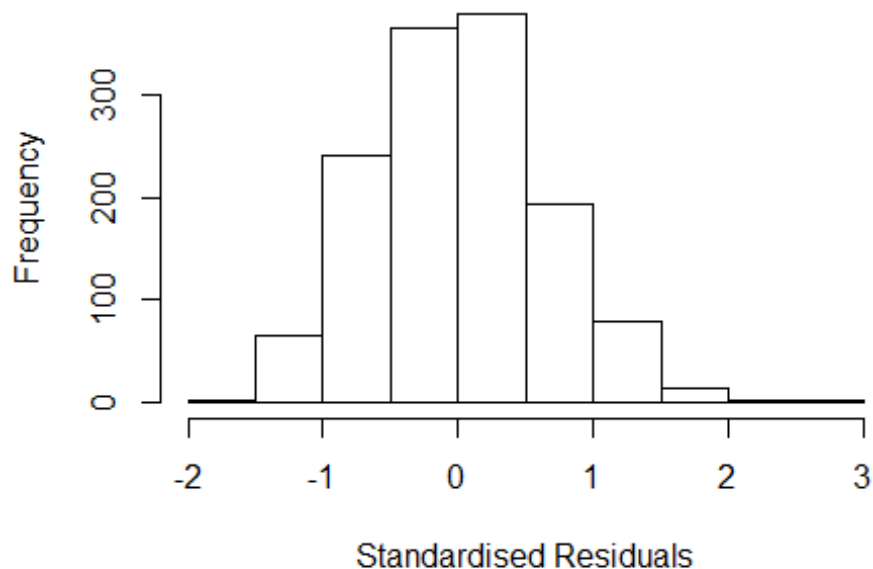


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.29767845991416"
## [1] "Male first author team size 2018 geometric mean: 2.99460369776735"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 740, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.58684452658227"
## [1] "Male last author team size 2018 geometric mean: 2.89332177446775"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.200 1      1.095
## LastAuthorFemale  1.204 1      1.097
## UniqueAuthors    1.249 4      1.028
## Year             1.332 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
## 3788 84857370722 4.1 2012      1705      3      2.702
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.51134 -0.45082 -0.00216  0.43320  2.70245
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.80131    0.10157   7.89 6.4e-15 ***
## FirstAuthorFemale1 -0.02524    0.05681  -0.44  0.657
## LastAuthorFemale1  0.01419    0.06078   0.23  0.815
## UniqueAuthors2    0.18746    0.04606   4.07 5.0e-05 ***
## UniqueAuthors3    0.36532    0.05369   6.80 1.5e-11 ***
## UniqueAuthors4    0.46492    0.08639   5.38 8.7e-08 ***
## UniqueAuthors5    0.38781    0.08005   4.84 1.4e-06 ***
## Year1997        -0.04221    0.13092  -0.32  0.747
## Year1998         0.11690    0.14592   0.80  0.423
## Year1999         0.00703    0.14800   0.05  0.962
```

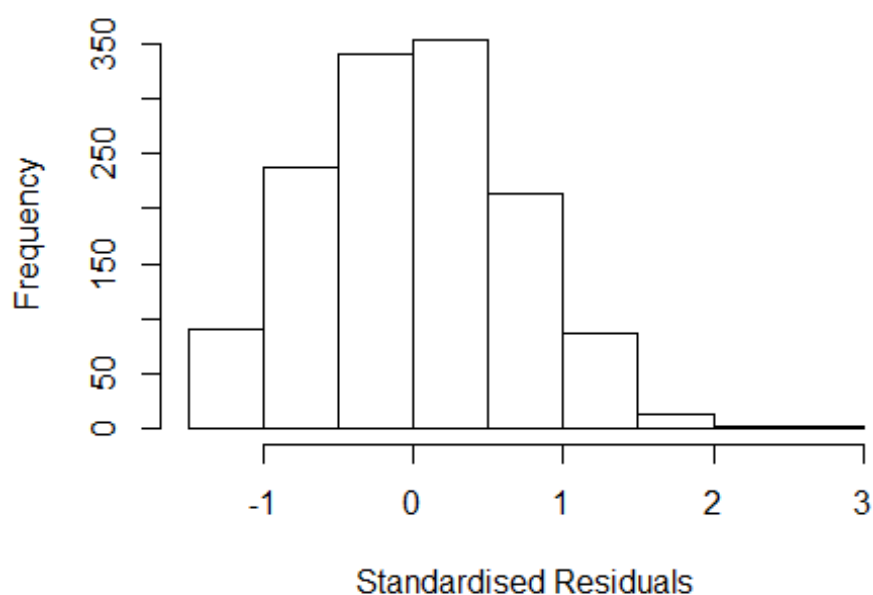


```

## Year2000      0.25548    0.12784    2.00    0.046 *
## Year2001      0.12692    0.13118    0.97    0.333
## Year2002      0.11546    0.14129    0.82    0.414
## Year2003      0.15495    0.13545    1.14    0.253
## Year2004      0.11119    0.13174    0.84    0.399
## Year2005      0.17072    0.11973    1.43    0.154
## Year2006      0.17472    0.12725    1.37    0.170
## Year2007      0.05835    0.11374    0.51    0.608
## Year2008     -0.04661    0.12712   -0.37    0.714
## Year2009      0.01029    0.11979    0.09    0.932
## Year2010      0.13917    0.12500    1.11    0.266
## Year2011      0.25860    0.11996    2.16    0.031 *
## Year2012      0.23092    0.12105    1.91    0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.079, Adjusted R-squared:  0.0635
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1222 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0553 0.8670 0.9510 0.9100 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      7.47e-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 1.183 1      1.088
## LastAuthorFemale  1.205 1      1.098
## Year              1.076 16      1.002

```

Residuals from first and last author

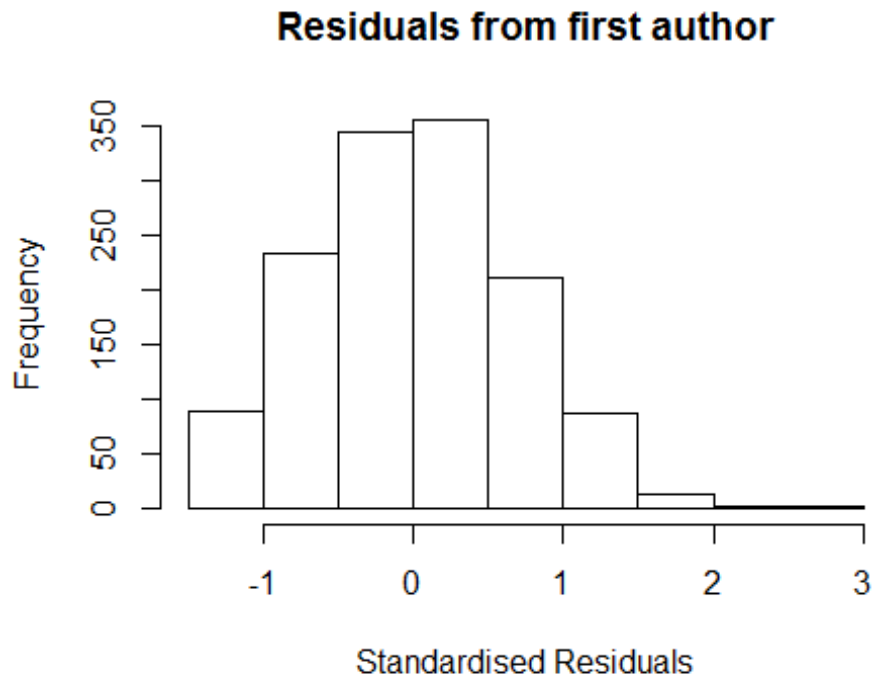


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3788 84857370722 4.1 2012      1705      3      2.817
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30702 -0.48437  0.00144  0.46560  2.81734
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90988    0.10016   9.08  <2e-16 ***
## FirstAuthorFemale1 -0.01080    0.05661  -0.19   0.8487
## LastAuthorFemale1  0.02436    0.06108   0.40   0.6901
## Year1997        -0.00564    0.13166  -0.04   0.9659
## Year1998         0.15832    0.14394   1.10   0.2716
## Year1999         0.03129    0.14612   0.21   0.8305
## Year2000         0.29333    0.12929   2.27   0.0234 *
## Year2001         0.16118    0.13261   1.22   0.2244
## Year2002         0.16342    0.14111   1.16   0.2470
## Year2003         0.18472    0.13551   1.36   0.1731
## Year2004         0.16262    0.13283   1.22   0.2211
## Year2005         0.26714    0.12024   2.22   0.0265 *
```

```

## Year2006          0.29764      0.12599      2.36      0.0183 *
## Year2007          0.12416      0.11487      1.08      0.2800
## Year2008          0.04889      0.12805      0.38      0.7027
## Year2009          0.14225      0.11847      1.20      0.2301
## Year2010          0.26857      0.12300      2.18      0.0292 *
## Year2011          0.37020      0.11853      3.12      0.0018 **
## Year2012          0.37278      0.11993      3.11      0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.0167
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1219 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0416 0.8560 0.9450 0.9100 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      7.47e-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 1.021 1      1.011
## Year              1.021 16      1.001

```

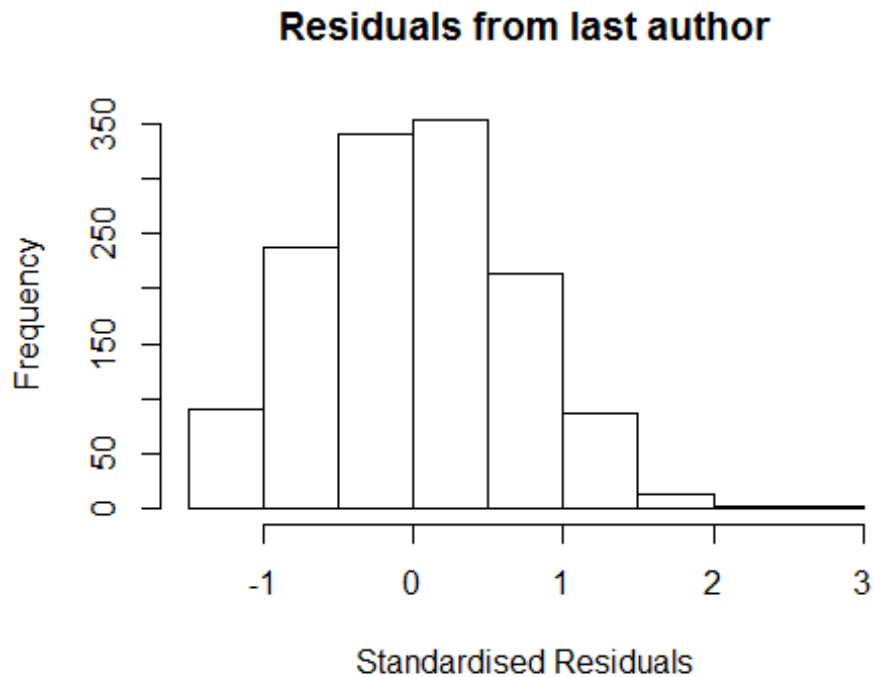


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3788 84857370722 4.1 2012      1705      3      2.817
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.285597 -0.484971  0.000916  0.461316  2.814403
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91062    0.10015   9.09  <2e-16 ***
## FirstAuthorFemale1 -0.00307    0.05257  -0.06   0.9534
## Year1997      -0.00473    0.13155  -0.04   0.9713
## Year1998       0.15784    0.14396   1.10   0.2731
## Year1999       0.03396    0.14595   0.23   0.8160
## Year2000       0.29484    0.12921   2.28   0.0227 *
## Year2001       0.16250    0.13252   1.23   0.2203
## Year2002       0.16462    0.14098   1.17   0.2432
## Year2003       0.18742    0.13516   1.39   0.1658
## Year2004       0.16711    0.13237   1.26   0.2070
## Year2005       0.26843    0.12000   2.24   0.0255 *
## Year2006       0.29797    0.12602   2.36   0.0182 *
```

```

## Year2007          0.12356      0.11482      1.08      0.2821
## Year2008          0.04904      0.12808      0.38      0.7019
## Year2009          0.14469      0.11822      1.22      0.2212
## Year2010          0.27115      0.12260      2.21      0.0272 *
## Year2011          0.37171      0.11828      3.14      0.0017 **
## Year2012          0.37498      0.11971      3.13      0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.0298, Adjusted R-squared:  0.0173
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1220 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0424 0.8560 0.9450 0.9100 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      7.47e-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 1.04 1      1.020
## Year      1.04 16      1.001

```



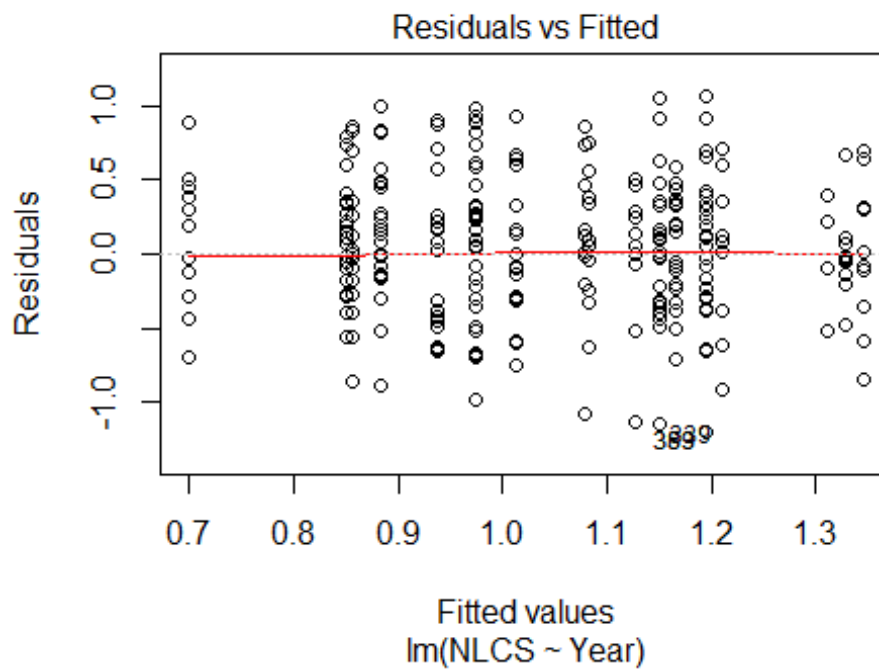
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3788 84857370722 4.1 2012      1705      3      2.817
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30235 -0.48357  0.00127  0.46353  2.81821
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90910    0.10005   9.09  <2e-16 ***
## LastAuthorFemale1 0.02055    0.05683   0.36  0.7177
## Year1997      -0.00567    0.13169  -0.04  0.9657
## Year1998       0.15787    0.14401   1.10  0.2732
## Year1999       0.03072    0.14629   0.21  0.8337
## Year2000       0.29305    0.12926   2.27  0.0235 *
## Year2001       0.16013    0.13270   1.21  0.2278
## Year2002       0.16285    0.14121   1.15  0.2490
## Year2003       0.18507    0.13556   1.37  0.1724
## Year2004       0.16280    0.13286   1.23  0.2207
## Year2005       0.26711    0.12024   2.22  0.0265 *
## Year2006       0.29740    0.12604   2.36  0.0184 *
```

```

## Year2007      0.12287      0.11461      1.07      0.2839
## Year2008      0.04791      0.12803      0.37      0.7083
## Year2009      0.14189      0.11853      1.20      0.2315
## Year2010      0.26828      0.12309      2.18      0.0295 *
## Year2011      0.37040      0.11851      3.13      0.0018 **
## Year2012      0.37269      0.11992      3.11      0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.673
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.0174
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1222 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0407 0.8560 0.9450 0.9100 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      7.47e-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: 1338"
## [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
##    6   17   13   11   25   32   11   20   15   13   27   22   41   24   28
## 2011 2012
##   33   40
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4   10   10   10   18   23    9   13   12   11   25   19   36   18   25
## 2011 2012

```

```
## 25 29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 9 10 8 17 21 8 13 12 11 25 18 34 16 25
## 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 = 16, df = 16, p-value = 0.5
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.087, df = 1, p-value = 0.8

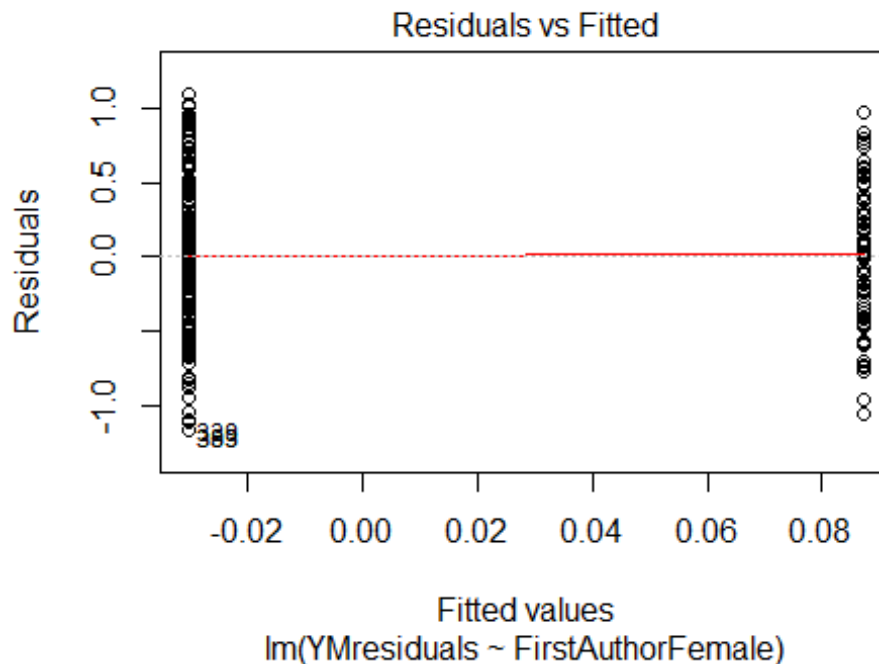
## [1] "Female first author team size 2018 geometric mean: 3.11214924268908"
## [1] "Male first author team size 2018 geometric mean: 2.72659788579485"

## 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 = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.07994615143797"
## [1] "Male last author team size 2018 geometric mean: 2.77153265489505"

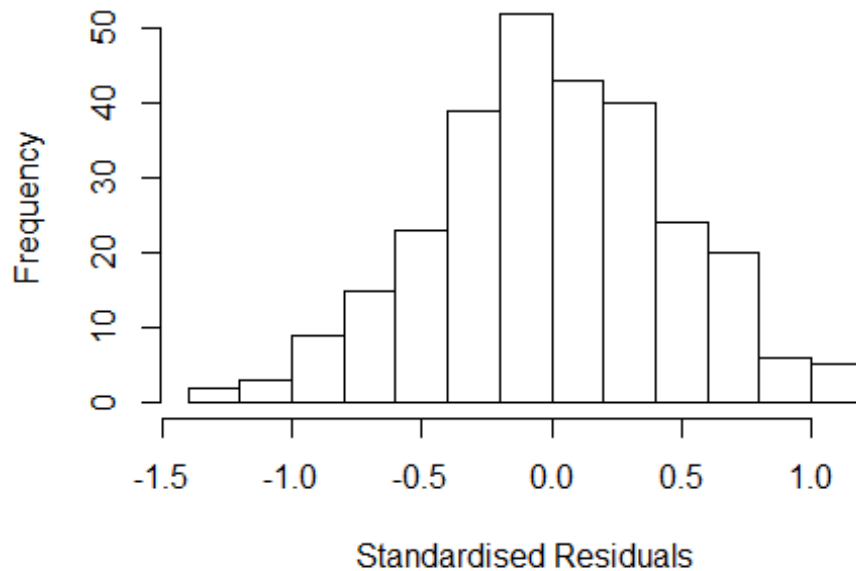
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.6
## 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	1.616	1	1.271
LastAuthorFemale	1.607	1	1.268
UniqueAuthors	2.389	4	1.115
Year	3.049	16	1.035

Residuals from first and last author and team size



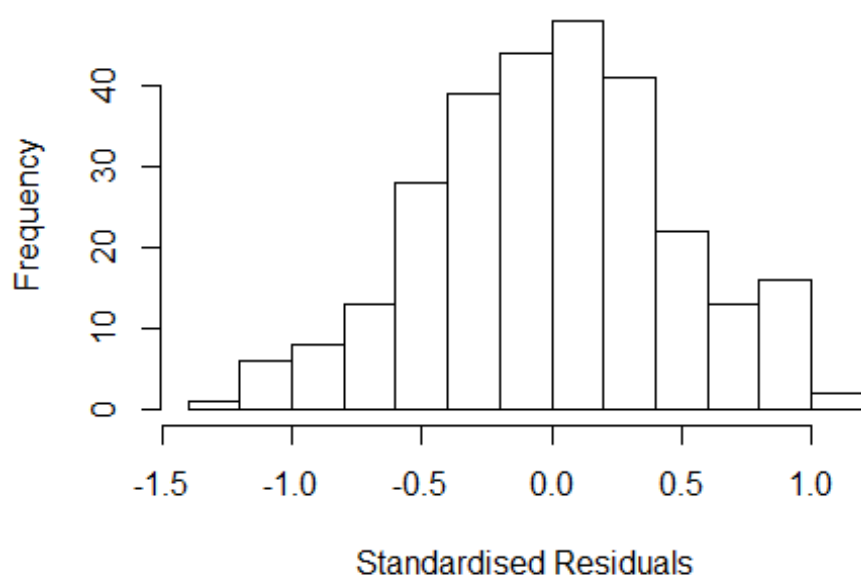
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26000 -0.31355 -0.00299 0.28673 1.13026
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23159 0.24150 5.10 6.6e-07 ***
## FirstAuthorFemale1 0.06154 0.08036 0.77 0.444
## LastAuthorFemale1 0.07992 0.08171 0.98 0.329
## UniqueAuthors2 0.16327 0.07205 2.27 0.024 *
## UniqueAuthors3 0.17450 0.08627 2.02 0.044 *
## UniqueAuthors4 0.26139 0.28331 0.92 0.357
## UniqueAuthors5 0.18922 0.20725 0.91 0.362
## Year1997 -0.13486 0.28359 -0.48 0.635
## Year1998 -0.20430 0.32943 -0.62 0.536
## Year1999 0.13466 0.27918 0.48 0.630
```

```

## Year2000      -0.29961    0.26773   -1.12    0.264
## Year2001      -0.16649    0.25660   -0.65    0.517
## Year2002       0.04335    0.30422    0.14    0.887
## Year2003      -0.71445    0.28022   -2.55    0.011 *
## Year2004      -0.28223    0.28522   -0.99    0.323
## Year2005      -0.00642    0.25814   -0.02    0.980
## Year2006      -0.45420    0.25365   -1.79    0.075 .
## Year2007      -0.50883    0.25538   -1.99    0.047 *
## Year2008      -0.39985    0.26797   -1.49    0.137
## Year2009      -0.43005    0.27477   -1.57    0.119
## Year2010      -0.21097    0.25732   -0.82    0.413
## Year2011      -0.13032    0.26493   -0.49    0.623
## Year2012      -0.47466    0.25989   -1.83    0.069 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.171, Adjusted R-squared:  0.0999
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 256 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.486  0.890  0.954  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      3.56e-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 1.673 1      1.294
## LastAuthorFemale  1.621 1      1.273
## Year              1.441 16      1.011

```

Residuals from first and last author



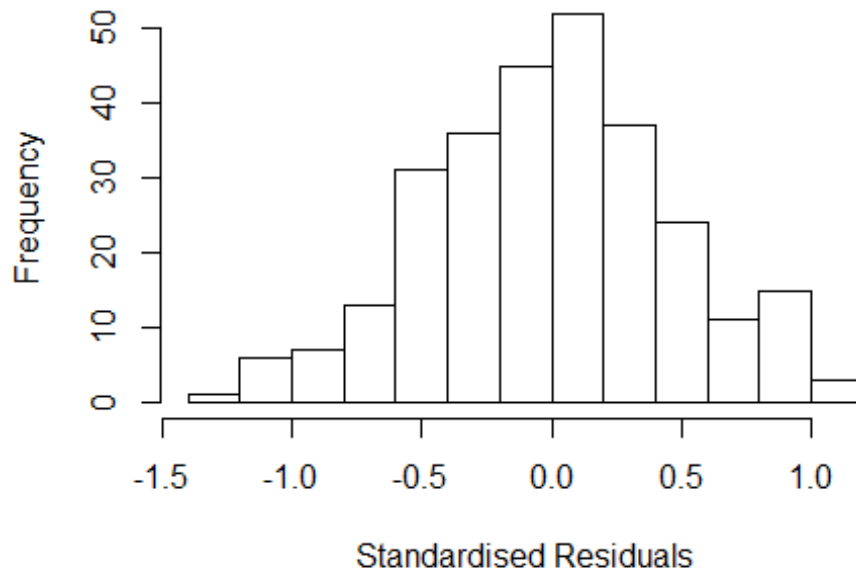
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.224445 -0.307582 0.000724 0.299417 1.032555
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31635 0.18425 7.14 9e-12 ***
## FirstAuthorFemale1 0.07563 0.08362 0.90 0.3666
## LastAuthorFemale1 0.05878 0.08446 0.70 0.4870
## Year1997 -0.17901 0.23389 -0.77 0.4448
## Year1998 -0.18956 0.28719 -0.66 0.5098
## Year1999 0.12410 0.24350 0.51 0.6107
## Year2000 -0.31077 0.21801 -1.43 0.1552
## Year2001 -0.14485 0.20371 -0.71 0.4777
## Year2002 0.00475 0.25735 0.02 0.9853
## Year2003 -0.68325 0.23605 -2.89 0.0041 **
## Year2004 -0.26335 0.23275 -1.13 0.2589
## Year2005 -0.02726 0.20142 -0.14 0.8924
```

```

## Year2006          -0.49019      0.20251    -2.42    0.0162 *
## Year2007          -0.53440      0.20673    -2.59    0.0103 *
## Year2008          -0.37377      0.22118    -1.69    0.0922 .
## Year2009          -0.44222      0.22807    -1.94    0.0536 .
## Year2010          -0.20892      0.20754    -1.01    0.3150
## Year2011          -0.09190      0.21564    -0.43    0.6703
## Year2012          -0.48949      0.21254    -2.30    0.0221 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.146, Adjusted R-squared:  0.0878
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 256 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.514  0.878   0.959   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      3.56e-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 1.227 1      1.107
## Year              1.227 16      1.006

```

Residuals from first author



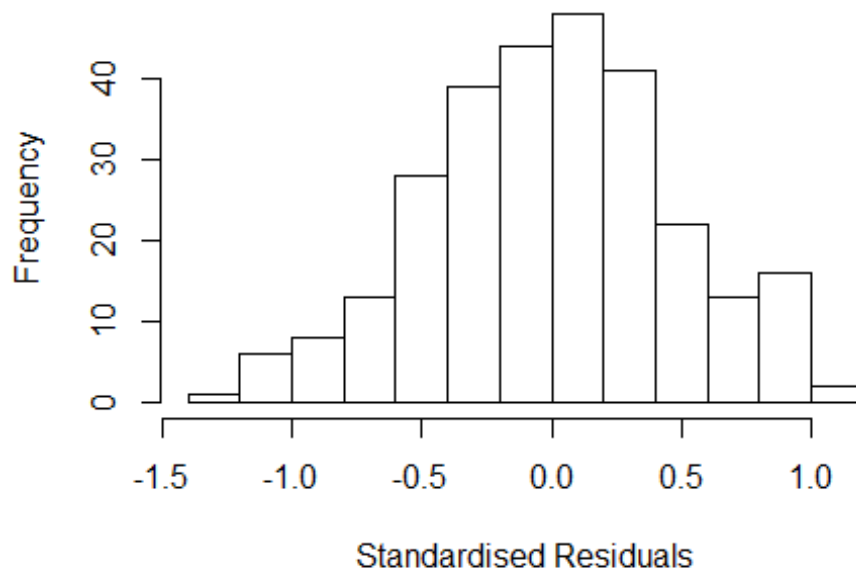
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23065 -0.30317 0.00482 0.29242 1.04198
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31642 0.18438 7.14 9.1e-12 ***
## FirstAuthorFemale1 0.10410 0.07132 1.46 0.1456
## Year1997 -0.16677 0.23272 -0.72 0.4743
## Year1998 -0.19785 0.28746 -0.69 0.4919
## Year1999 0.13514 0.24647 0.55 0.5840
## Year2000 -0.30671 0.21778 -1.41 0.1602
## Year2001 -0.13828 0.20287 -0.68 0.4961
## Year2002 0.00508 0.25746 0.02 0.9843
## Year2003 -0.67131 0.23438 -2.86 0.0045 **
## Year2004 -0.26025 0.23259 -1.12 0.2642
## Year2005 -0.02434 0.20128 -0.12 0.9039
## Year2006 -0.48822 0.20281 -2.41 0.0168 *
```

```

## Year2007          -0.53002    0.20666   -2.56   0.0109 *
## Year2008          -0.36684    0.22090   -1.66   0.0980 .
## Year2009          -0.43924    0.23020   -1.91   0.0575 .
## Year2010          -0.19479    0.20566   -0.95   0.3444
## Year2011          -0.08577    0.21631   -0.40   0.6920
## Year2012          -0.48540    0.21248   -2.28   0.0231 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.145, Adjusted R-squared:  0.0902
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 254 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.506  0.880  0.953  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      3.56e-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 1.178 1          1.085
## Year              1.178 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
## -1.24166 -0.30894 0.00662 0.31106 1.01534
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31621 0.18402 7.15 8.4e-12 ***
## LastAuthorFemale1 0.09945 0.07186 1.38 0.1676
## Year1997 -0.16706 0.23724 -0.70 0.4819
## Year1998 -0.16715 0.28000 -0.60 0.5510
## Year1999 0.12470 0.24180 0.52 0.6065
## Year2000 -0.30059 0.21889 -1.37 0.1708
## Year2001 -0.13854 0.20323 -0.68 0.4960
## Year2002 0.00414 0.25713 0.02 0.9872
## Year2003 -0.66452 0.23779 -2.79 0.0056 **
## Year2004 -0.25312 0.23051 -1.10 0.2732
## Year2005 -0.01252 0.20025 -0.06 0.9502
## Year2006 -0.48283 0.20262 -2.38 0.0179 *
```



```

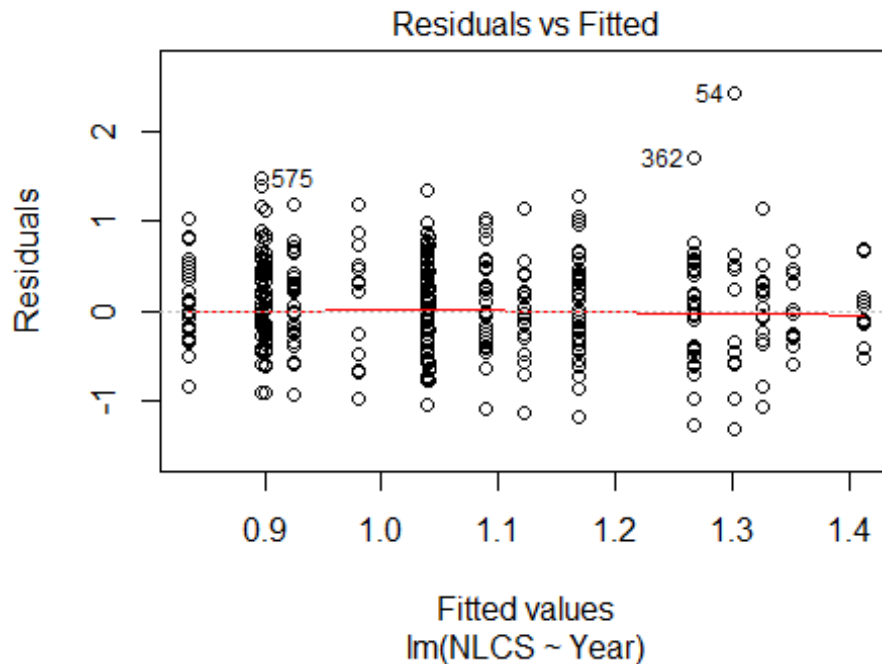
## Year2007          -0.52925      0.20732    -2.55    0.0113 *
## Year2008          -0.36093      0.21958    -1.64    0.1014
## Year2009          -0.43099      0.22613    -1.91    0.0578 .
## Year2010          -0.20627      0.20834    -0.99    0.3230
## Year2011          -0.07455      0.21401    -0.35    0.7279
## Year2012          -0.47227      0.21130    -2.24    0.0263 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.143, Adjusted R-squared:  0.0877
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.511  0.881  0.958  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      3.56e-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: 281"
## [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]"
##
## 2006 2008 2009 2010 2011 2012
##    3    1    1    1    3    1
##
## 2006 2008 2009 2010 2011 2012
##    3    1    0    0    3    1
##
## 2006 2008 2009 2010 2011 2012
##    3    0    0    0    3    1

```

```

## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2"
## [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: 7"
## [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
## 25 19 21 33 26 24 23 21 37 25 49 58 72 72 84
## 2011 2012
## 60 62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 11 15 19 15 11 15 11 31 18 32 38 52 54 60
## 2011 2012
## 35 39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 11 14 17 14 10 14 8 27 16 25 33 47 48 49
## 2011 2012
## 28 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 = 23, df = 16, p-value = 0.1

```



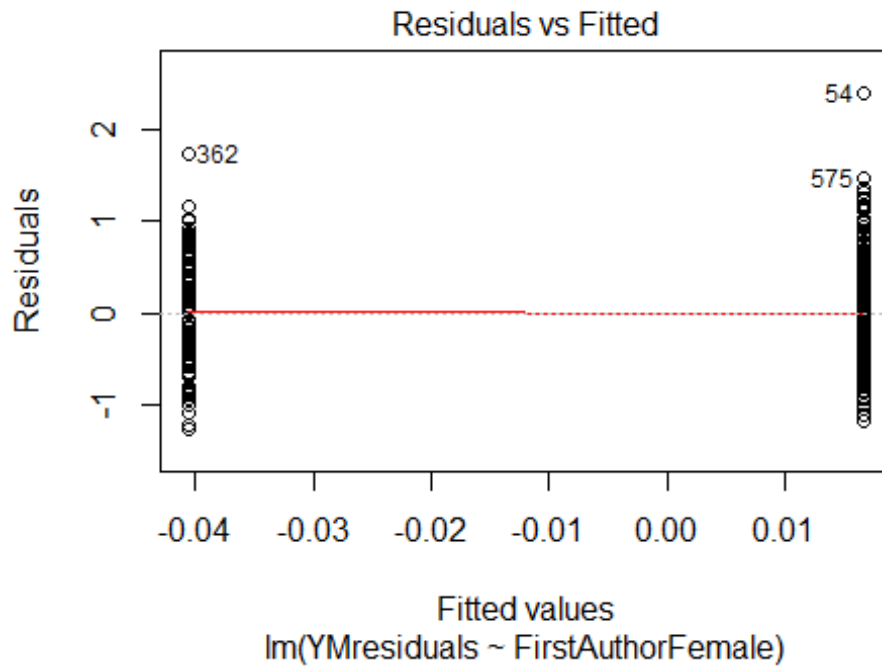
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 2.84731468735758"
## [1] "Male first author team size 2018 geometric mean: 2.63646233809154"

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

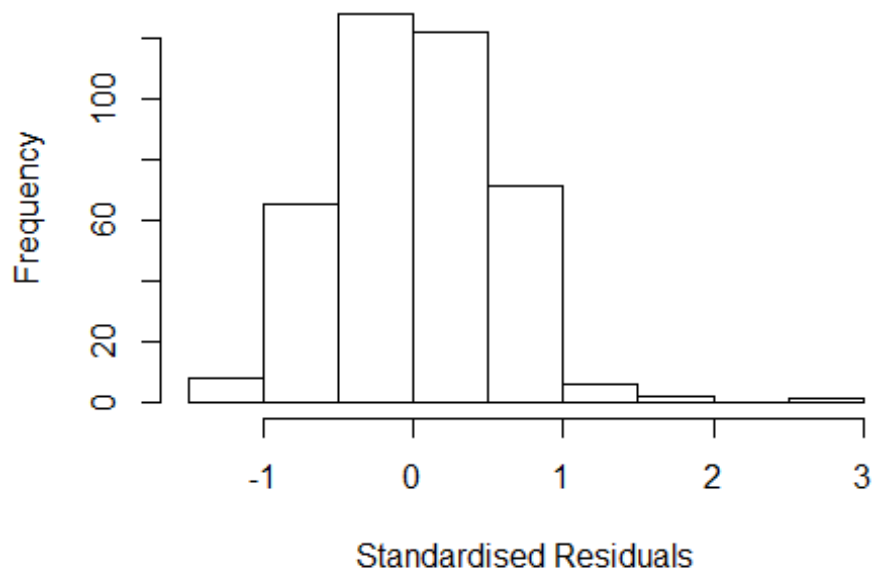
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 150, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.99627555782726"
## [1] "Male last author team size 2018 geometric mean: 2.58395824969125"

## 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 = 0.6
## 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 1.800 1          1.342
## LastAuthorFemale  1.747 1          1.322
## UniqueAuthors    1.988 4          1.090
## Year             2.559 16          1.030
```

Residuals from first and last author and team size



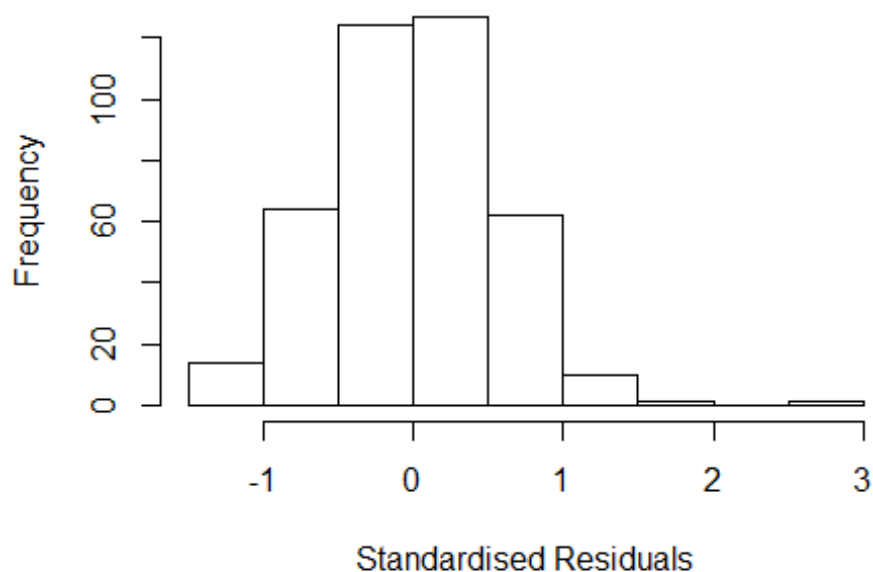
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 54 0031482189 3.717 1997    1705      4      2.715
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.25019 -0.35334  0.00203  0.39127  2.71492
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90223    0.23838   3.78 0.00018 ***
## FirstAuthorFemale1 0.03177    0.08243   0.39 0.70016
## LastAuthorFemale1 -0.08119    0.08003  -1.01 0.31099
## UniqueAuthors2    0.13081    0.06828   1.92 0.05612 .
## UniqueAuthors3    0.25986    0.07613   3.41 0.00071 ***
## UniqueAuthors4    0.20978    0.19074   1.10 0.27210
## UniqueAuthors5    0.46024    0.29038   1.58 0.11381
## Year1997         0.09985    0.35188   0.28 0.77676
## Year1998         0.19617    0.29431   0.67 0.50547
## Year1999         0.15269    0.27635   0.55 0.58092
```

```

## Year2000      0.12116      0.26629      0.46  0.64937
## Year2001      0.30885      0.28175      1.10  0.27369
## Year2002      0.36762      0.27687      1.33  0.18505
## Year2003      0.35558      0.28881      1.23  0.21902
## Year2004      0.05633      0.26363      0.21  0.83092
## Year2005      0.07489      0.28772      0.26  0.79480
## Year2006      0.37440      0.26486      1.41  0.15831
## Year2007     -0.12379      0.25809     -0.48  0.63176
## Year2008      0.14493      0.25611      0.57  0.57180
## Year2009     -0.09291      0.25967     -0.36  0.72069
## Year2010     -0.00219      0.26367     -0.01  0.99337
## Year2011     -0.16814      0.25849     -0.65  0.51578
## Year2012     -0.13552      0.26209     -0.52  0.60541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.118, Adjusted R-squared:  0.0672
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 25 is an outlier with |weight| <= 4.4e-05 ( < 0.00025);
## 28 weights are ~= 1. The remaining 374 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.356  0.868  0.955  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           2.48e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.774 1 1.332
## LastAuthorFemale 1.735 1 1.317
## Year 1.387 16 1.010

```

Residuals from first and last author



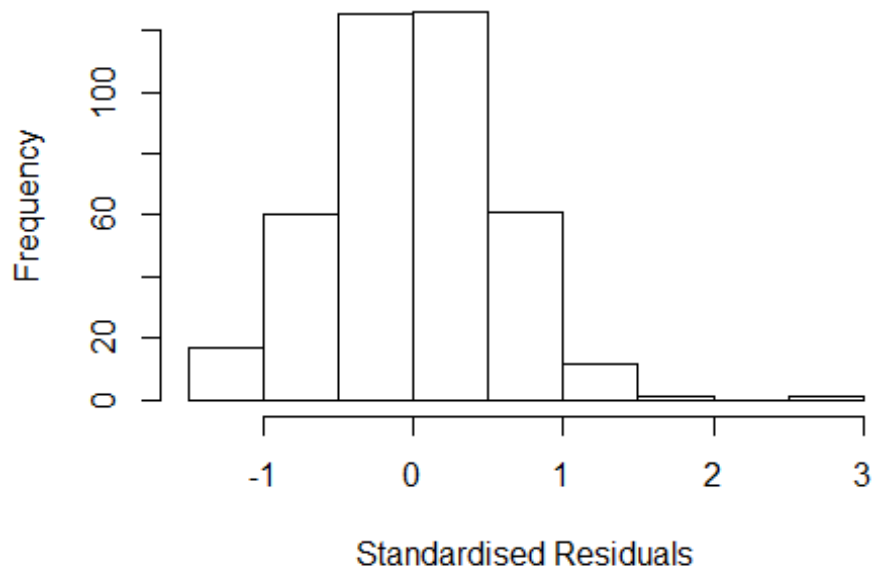
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 54 0031482189 3.717 1997      1705      4      2.621
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.19084 -0.35574 -0.00398  0.41591  2.62133
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93521    0.22685   4.12 4.6e-05 ***
## FirstAuthorFemale1 0.03737    0.08465   0.44  0.659
## LastAuthorFemale1 -0.09862    0.08187  -1.20  0.229
## Year1997         0.16046    0.32950   0.49  0.627
## Year1998         0.25562    0.28301   0.90  0.367
## Year1999         0.18403    0.26155   0.70  0.482
## Year2000         0.15484    0.25941   0.60  0.551
## Year2001         0.38377    0.25268   1.52  0.130
## Year2002         0.40348    0.26886   1.50  0.134
## Year2003         0.40410    0.27307   1.48  0.140
## Year2004         0.10521    0.24899   0.42  0.673
## Year2005         0.19026    0.26295   0.72  0.470
```

```

## Year2006          0.45187      0.25056      1.80      0.072 .
## Year2007          -0.05010      0.24397     -0.21      0.837
## Year2008           0.23786      0.24278      0.98      0.328
## Year2009          -0.00938      0.24698     -0.04      0.970
## Year2010           0.09284      0.25016      0.37      0.711
## Year2011          -0.12958      0.24776     -0.52      0.601
## Year2012          -0.06032      0.24986     -0.24      0.809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0865, Adjusted R-squared:  0.0437
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 375 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0077 0.8750 0.9530 0.9160 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      2.48e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.161 1      1.078
## Year              1.161 16      1.005

```


Residuals from first author



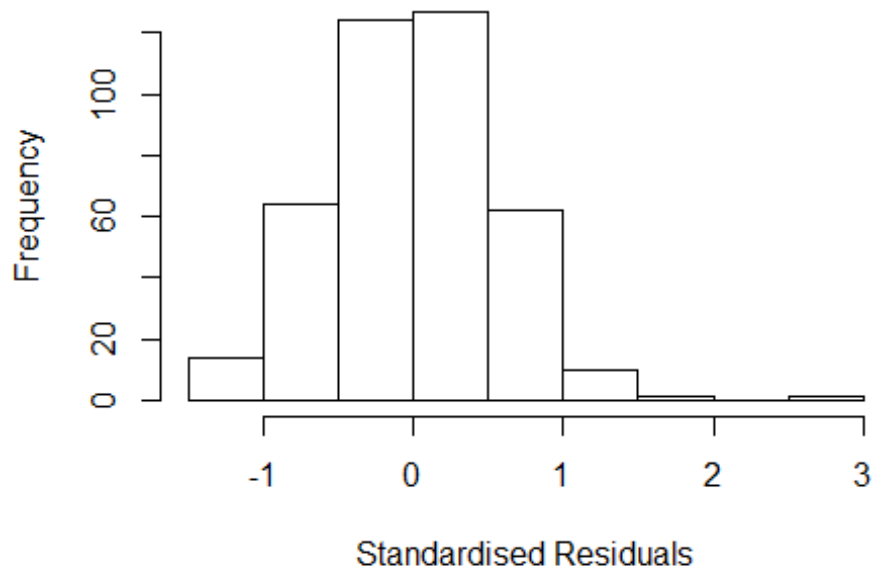
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 54 0031482189 3.717 1997      1705      4      2.621
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18194 -0.37304 -0.00144  0.41211  2.65049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9260     0.2285   4.05 6.1e-05 ***
## FirstAuthorFemale1 -0.0144     0.0693  -0.21  0.836
## Year1997          0.1405     0.3345   0.42  0.675
## Year1998          0.2560     0.2792   0.92  0.360
## Year1999          0.1852     0.2650   0.70  0.485
## Year2000          0.1698     0.2620   0.65  0.517
## Year2001          0.4149     0.2524   1.64  0.101
## Year2002          0.4043     0.2694   1.50  0.134
## Year2003          0.4009     0.2797   1.43  0.153
## Year2004          0.0904     0.2486   0.36  0.716
## Year2005          0.1838     0.2634   0.70  0.486
## Year2006          0.4540     0.2534   1.79  0.074 .
```

```

## Year2007          -0.0629      0.2461   -0.26    0.798
## Year2008           0.2319      0.2441    0.95    0.343
## Year2009          -0.0214      0.2493   -0.09    0.932
## Year2010           0.0921      0.2507    0.37    0.714
## Year2011          -0.1380      0.2486   -0.56    0.579
## Year2012          -0.0630      0.2513   -0.25    0.802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0831, Adjusted R-squared:  0.0426
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 372 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0048 0.8770 0.9520 0.9150 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      2.48e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.133 1          1.065
## Year            1.133 16          1.004

```

Residuals from last author



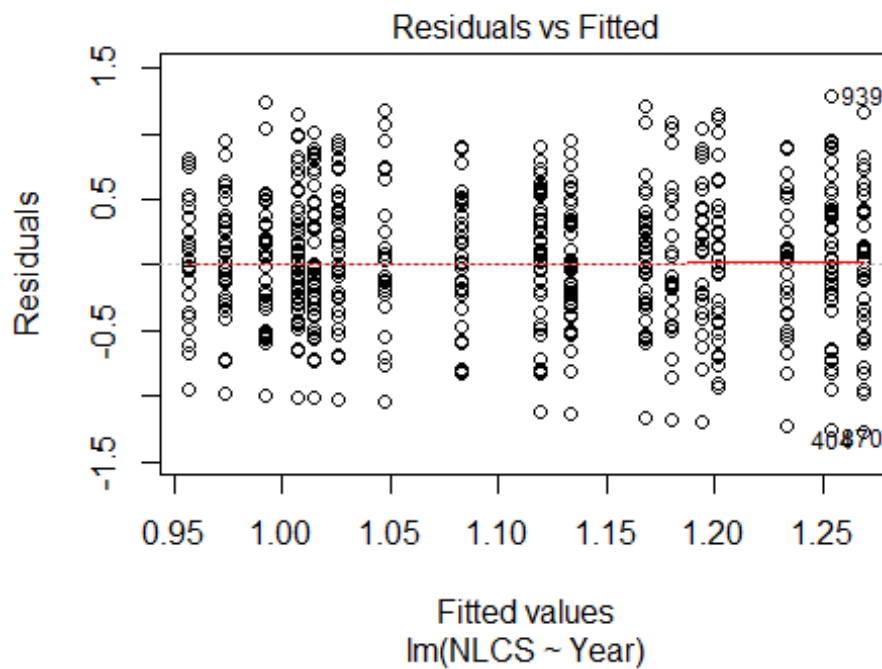
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 54 0031482189 3.717 1997    1705      4      2.621
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.190973 -0.357010 -0.000177  0.416002  2.626035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93850    0.23067   4.07 5.7e-05 ***
## LastAuthorFemale1 -0.07871    0.06609  -1.19  0.234
## Year1997        0.15246    0.32984   0.46  0.644
## Year1998        0.25247    0.28455   0.89  0.375
## Year1999        0.18563    0.26552   0.70  0.485
## Year2000        0.16005    0.26360   0.61  0.544
## Year2001        0.39634    0.25274   1.57  0.118
## Year2002        0.40372    0.27208   1.48  0.139
## Year2003        0.40559    0.27730   1.46  0.144
## Year2004        0.10561    0.25247   0.42  0.676
## Year2005        0.18893    0.26591   0.71  0.478
## Year2006        0.45239    0.25378   1.78  0.075 .
```

```

## Year2007          -0.04672      0.24786    -0.19      0.851
## Year2008           0.24091      0.24592      0.98      0.328
## Year2009          -0.00617      0.25036    -0.02      0.980
## Year2010           0.09640      0.25328      0.38      0.704
## Year2011          -0.13185      0.25138    -0.52      0.600
## Year2012          -0.05816      0.25338    -0.23      0.819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0859, Adjusted R-squared:  0.0455
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 376 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0078 0.8770 0.9510 0.9160 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      2.48e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 403"
## [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
## 46 65 47 61 55 73 57 49 49 47 78 71 79 70 76
## 2011 2012
## 70 84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 30 24 30 27 37 39 27 39 33 50 45 52 41 47
## 2011 2012

```

```
## 41 61
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 30 22 29 24 35 32 27 35 32 45 42 46 35 42
## 2011 2012
## 39 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 = 11, df = 16, p-value = 0.8
```



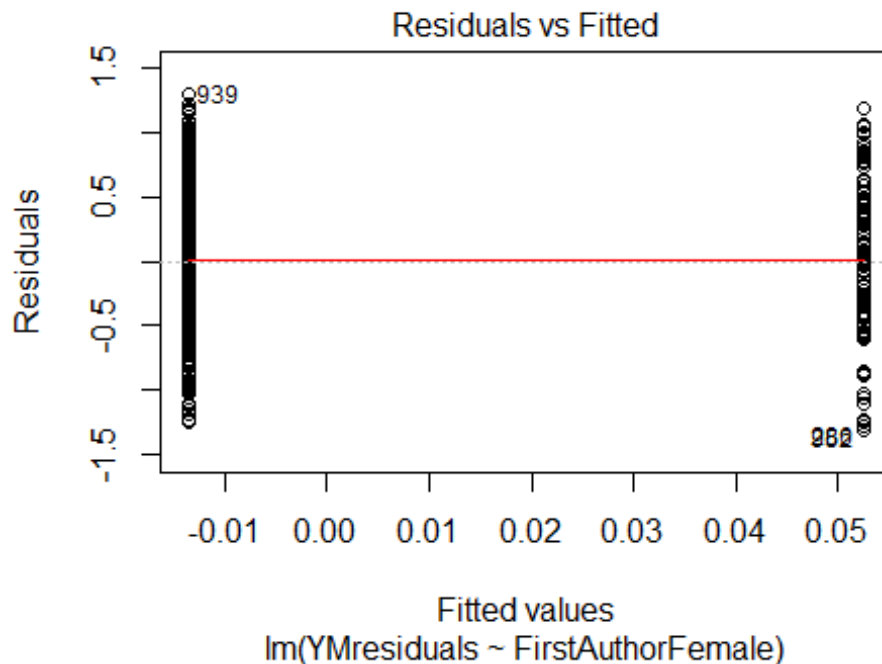
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.059, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 2.15522885553986"
## [1] "Male first author team size 2018 geometric mean: 2.72228333530224"

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

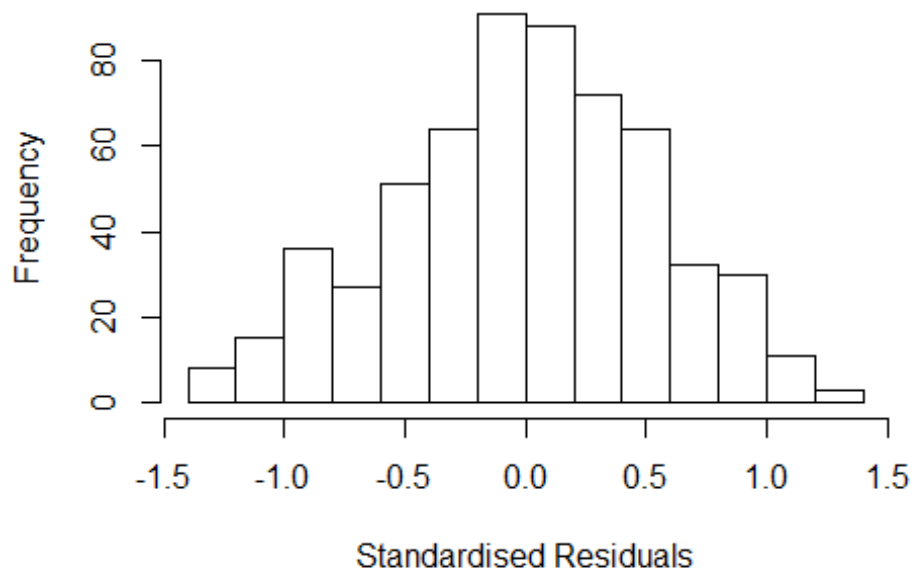
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.10397901101729"
## [1] "Male last author team size 2018 geometric mean: 2.65167177895802"

## 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 = 0.5
## 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 1.946 1      1.395
## LastAuthorFemale  1.886 1      1.373
## UniqueAuthors    1.924 4      1.085
## Year              2.427 16     1.028
```

Residuals from first and last author and team size



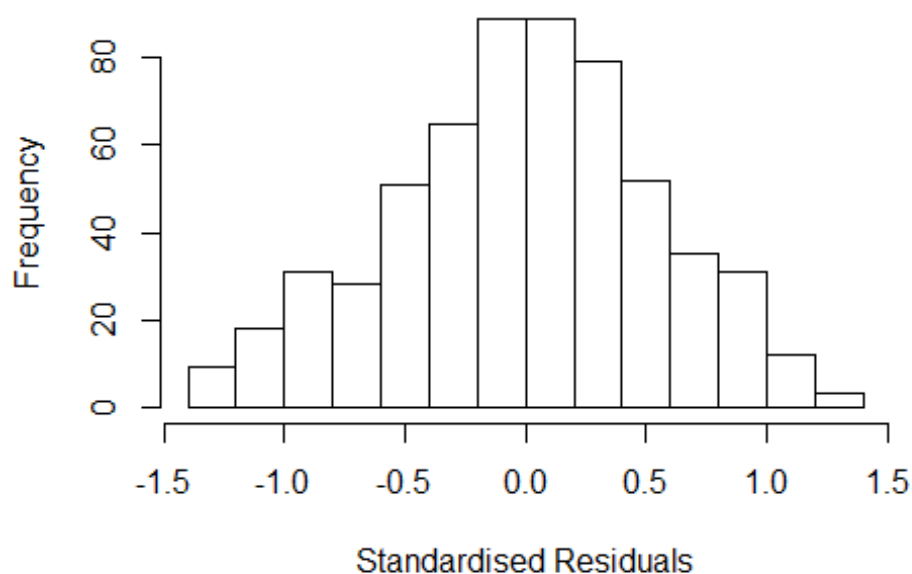
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33525 -0.35439  0.00186  0.37266  1.31487
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90740    0.09465   9.59   <2e-16 ***
## FirstAuthorFemale1 0.06316    0.07459   0.85   0.397
## LastAuthorFemale1 0.01967    0.06821   0.29   0.773
## UniqueAuthors2    0.05727    0.05610   1.02   0.308
## UniqueAuthors3    0.04796    0.07809   0.61   0.539
## UniqueAuthors4    0.20647    0.09883   2.09   0.037 *
## UniqueAuthors5    0.18375    0.21031   0.87   0.383
## Year1997          0.09008    0.14255   0.63   0.528
## Year1998          0.25460    0.16167   1.57   0.116
## Year1999          0.25693    0.15058   1.71   0.088 .
```

```

## Year2000      0.25915      0.14761      1.76      0.080 .
## Year2001      0.34380      0.13671      2.51      0.012 *
## Year2002      0.18928      0.12707      1.49      0.137
## Year2003      0.26768      0.16534      1.62      0.106
## Year2004     -0.00876      0.12353     -0.07      0.944
## Year2005      0.03491      0.12858      0.27      0.786
## Year2006      0.19008      0.12922      1.47      0.142
## Year2007     -0.00995      0.13267     -0.07      0.940
## Year2008      0.31673      0.13091      2.42      0.016 *
## Year2009      0.22273      0.12682      1.76      0.080 .
## Year2010      0.12021      0.12632      0.95      0.342
## Year2011      0.07785      0.14259      0.55      0.585
## Year2012     -0.00781      0.12023     -0.06      0.948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.052, Adjusted R-squared:  0.0153
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 529 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.546  0.858  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           1.69e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.875 1           1.369
## LastAuthorFemale  1.851 1           1.360
## Year              1.355 16           1.010

```


Residuals from first and last author



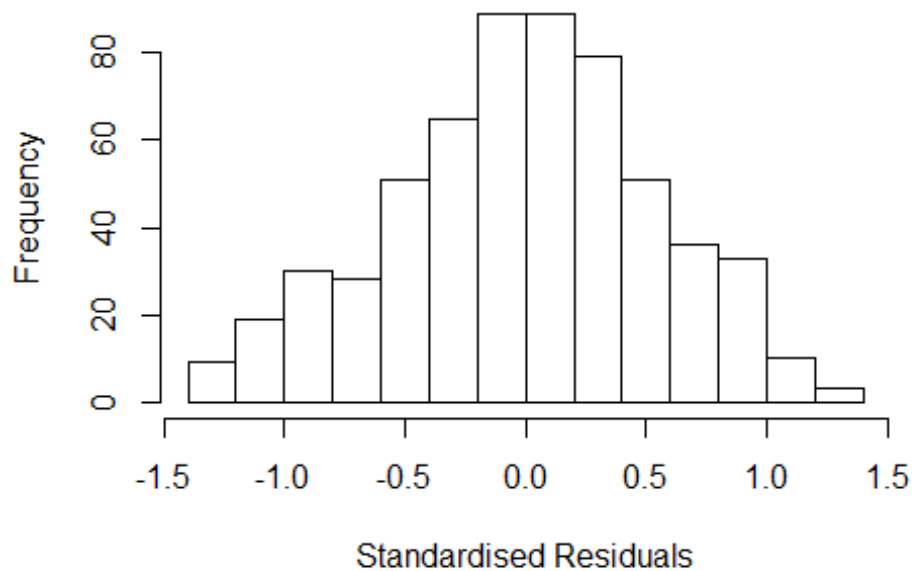
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3303 -0.3613  0.0141  0.3647  1.2726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93031    0.09190   10.12  <2e-16 ***
## FirstAuthorFemale1 0.06386    0.07475    0.85   0.393
## LastAuthorFemale1 0.02243    0.06850    0.33   0.743
## Year1997         0.08864    0.14053    0.63   0.528
## Year1998         0.25681    0.15937    1.61   0.108
## Year1999         0.26634    0.15150    1.76   0.079 .
## Year2000         0.26290    0.14699    1.79   0.074 .
## Year2001         0.34706    0.13646    2.54   0.011 *
## Year2002         0.19287    0.12645    1.53   0.128
## Year2003         0.28950    0.16446    1.76   0.079 .
## Year2004         0.00296    0.12573    0.02   0.981
## Year2005         0.05222    0.12856    0.41   0.685
```

```

## Year2006          0.19865      0.13087      1.52      0.130
## Year2007          0.01800      0.13084      0.14      0.891
## Year2008          0.33611      0.13024      2.58      0.010 *
## Year2009          0.25412      0.12595      2.02      0.044 *
## Year2010          0.14892      0.12057      1.24      0.217
## Year2011          0.09850      0.14413      0.68      0.495
## Year2012          0.02297      0.11819      0.19      0.846
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0456, Adjusted R-squared:  0.0156
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 538 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.547  0.864  0.953  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      1.69e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.212 1          1.101
## Year              1.212 16          1.006

```

Residuals from first author



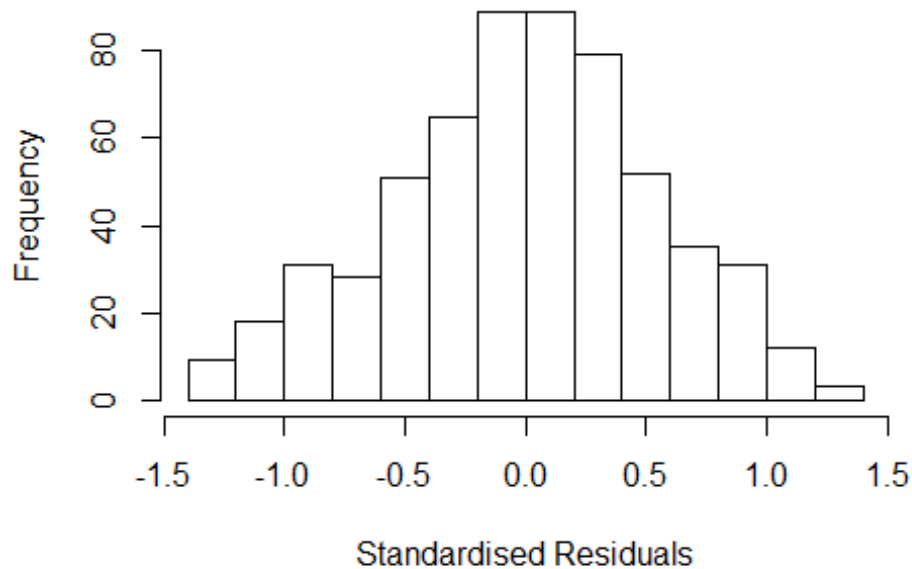
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3438 -0.3610 0.0154 0.3686 1.2703
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93187 0.09181 10.15 <2e-16 ***
## FirstAuthorFemale1 0.07519 0.06125 1.23 0.220
## Year1997 0.08814 0.14048 0.63 0.531
## Year1998 0.25828 0.15971 1.62 0.106
## Year1999 0.26549 0.15162 1.75 0.080 .
## Year2000 0.26081 0.14702 1.77 0.077 .
## Year2001 0.34853 0.13665 2.55 0.011 *
## Year2002 0.19214 0.12664 1.52 0.130
## Year2003 0.29049 0.16411 1.77 0.077 .
## Year2004 0.00213 0.12565 0.02 0.987
## Year2005 0.05103 0.12855 0.40 0.692
## Year2006 0.20205 0.12976 1.56 0.120
```

```

## Year2007          0.02002    0.13059    0.15    0.878
## Year2008          0.33679    0.13036    2.58    0.010 *
## Year2009          0.25640    0.12515    2.05    0.041 *
## Year2010          0.14935    0.12064    1.24    0.216
## Year2011          0.09837    0.14429    0.68    0.496
## Year2012          0.02266    0.11831    0.19    0.848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0172
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 534 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.538  0.862  0.953  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      1.69e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.215 1      1.102
## Year              1.215 16      1.006

```

Residuals from last author



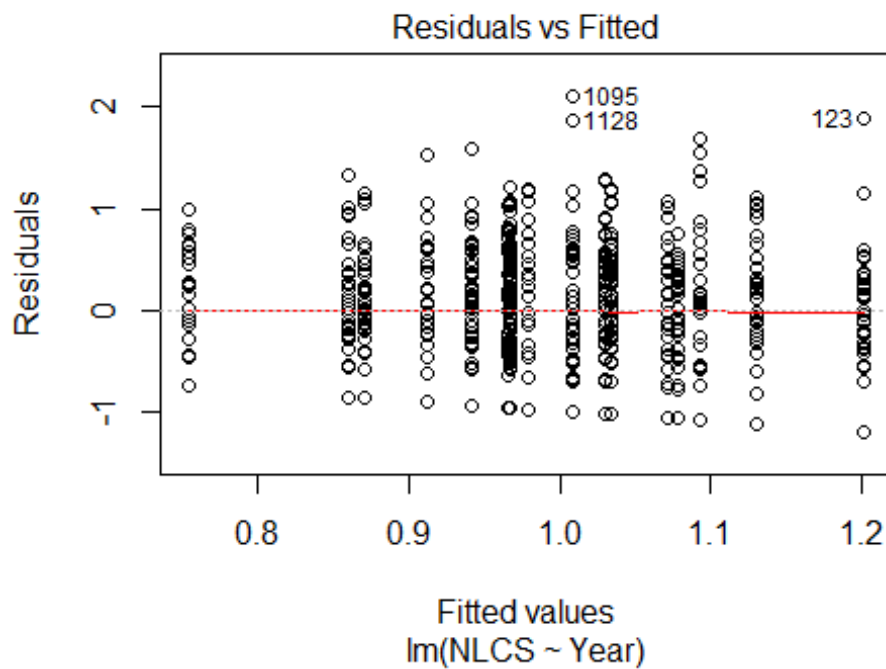
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2810 -0.3616 0.0144 0.3614 1.2656
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9345 0.0919 10.17 <2e-16 ***
## LastAuthorFemale1 0.0564 0.0555 1.02 0.3104
## Year1997 0.0910 0.1409 0.65 0.5186
## Year1998 0.2581 0.1602 1.61 0.1078
## Year1999 0.2627 0.1523 1.73 0.0850 .
## Year2000 0.2757 0.1469 1.88 0.0610 .
## Year2001 0.3465 0.1366 2.54 0.0114 *
## Year2002 0.1954 0.1256 1.56 0.1204
## Year2003 0.2848 0.1653 1.72 0.0855 .
## Year2004 0.0103 0.1266 0.08 0.9352
## Year2005 0.0523 0.1298 0.40 0.6872
## Year2006 0.2000 0.1303 1.53 0.1254
```

```

## Year2007          0.0222      0.1304      0.17      0.8651
## Year2008          0.3389      0.1307      2.59      0.0097 **
## Year2009          0.2539      0.1266      2.01      0.0453 *
## Year2010          0.1592      0.1190      1.34      0.1816
## Year2011          0.1008      0.1453      0.69      0.4882
## Year2012          0.0244      0.1189      0.21      0.8374
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0443, Adjusted R-squared:  0.016
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 539 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.574  0.859  0.954  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      1.69e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 592"
## [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
##   56   60   60   70   54   56   55   49   56   58   53   56   82   86   77
## 2011 2012
##   97   87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   38   38   31   31   35   35   35   39   37   35   33   53   57   55
## 2011 2012

```

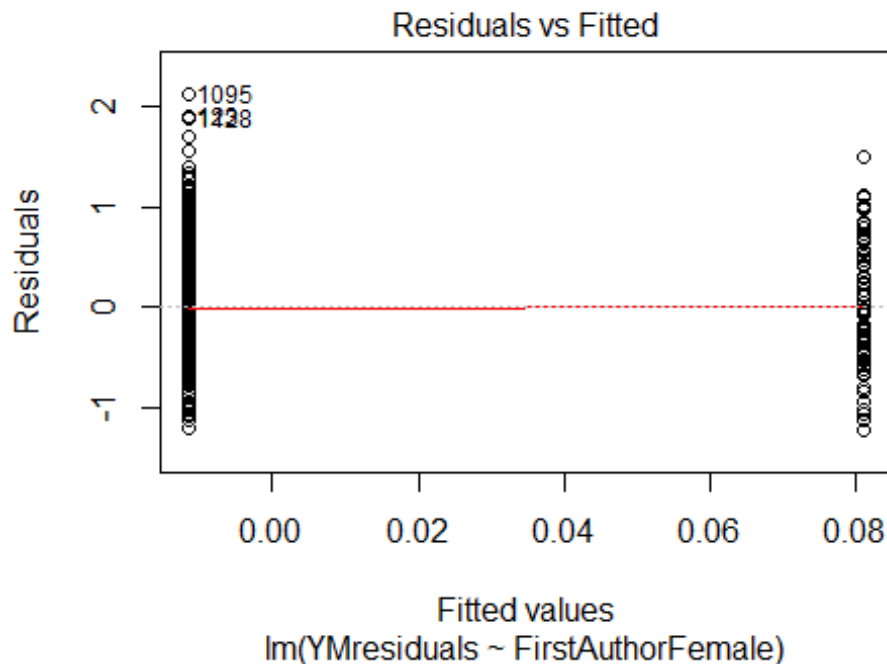
```
## 73 69
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 25 30 26 27 28 27 30 33 34 29 29 49 47 48
## 2011 2012
## 65 65
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```



```
##
## 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: 1.8881750225898"
## [1] "Male first author team size 2018 geometric mean: 1.82552990636525"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

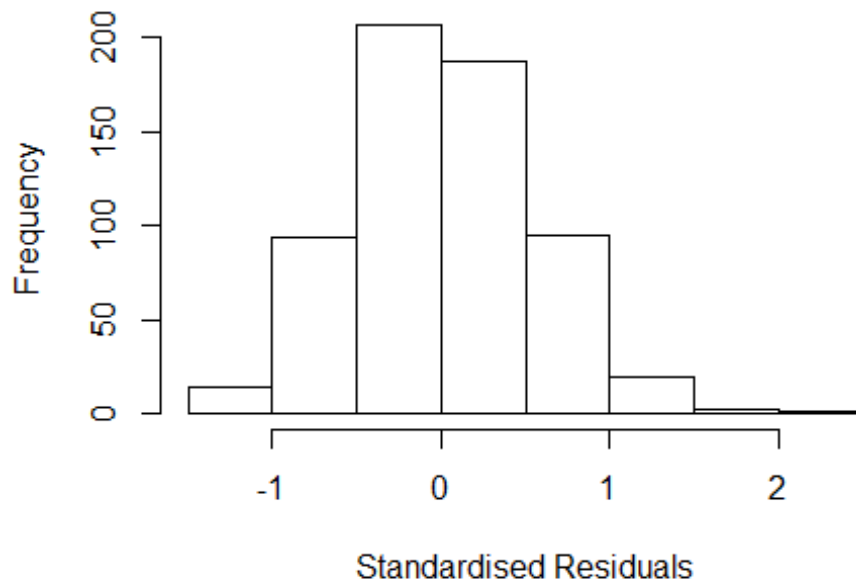
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.45950948584936"
## [1] "Male last author team size 2018 geometric mean: 1.76312403618046"

## 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 = 0.2
## 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 1.352 1          1.163
## LastAuthorFemale  1.285 1          1.134
## UniqueAuthors    2.087 4          1.096
## Year              2.015 16         1.022
```


Residuals from first and last author and team size



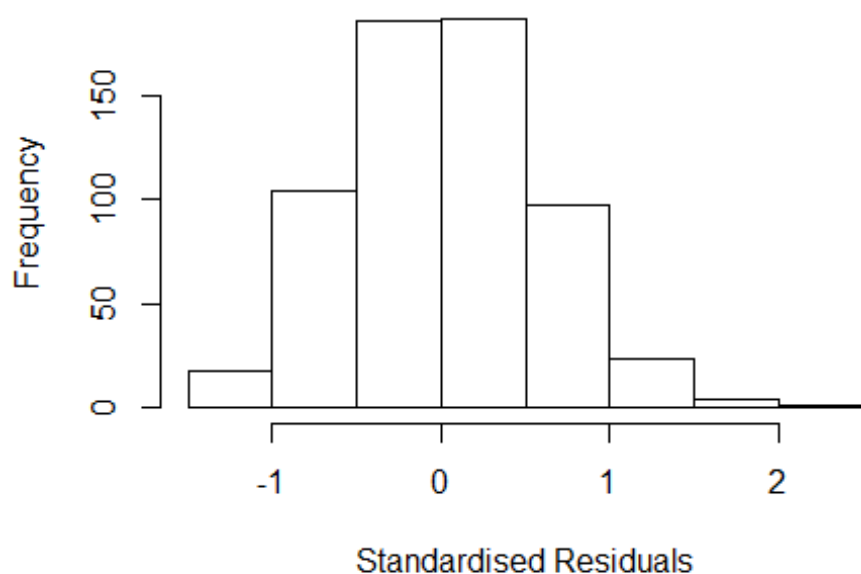
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26833 -0.36932 -0.00698 0.38502 2.11255
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94176 0.16253 5.79 1.1e-08 ***
## FirstAuthorFemale1 0.08840 0.09331 0.95 0.34385
## LastAuthorFemale1 0.00744 0.07884 0.09 0.92486
## UniqueAuthors2 0.25446 0.05358 4.75 2.6e-06 ***
## UniqueAuthors3 0.27207 0.07444 3.65 0.00028 ***
## UniqueAuthors4 0.48817 0.17550 2.78 0.00558 **
## UniqueAuthors5 -0.14261 0.12180 -1.17 0.24212
## Year1997 0.07791 0.19097 0.41 0.68343
## Year1998 -0.08378 0.19946 -0.42 0.67462
## Year1999 -0.13035 0.20555 -0.63 0.52621
```

```

## Year2000      -0.08994      0.20804      -0.43      0.66565
## Year2001      0.05450      0.20441      0.27      0.78987
## Year2002     -0.03815      0.18650     -0.20      0.83800
## Year2003     -0.07988      0.18198     -0.44      0.66087
## Year2004     -0.26231      0.19329     -1.36      0.17527
## Year2005     -0.29386      0.19231     -1.53      0.12704
## Year2006     -0.22615      0.20684     -1.09      0.27467
## Year2007     -0.34015      0.19304     -1.76      0.07857 .
## Year2008     -0.19378      0.18661     -1.04      0.29950
## Year2009     -0.08721      0.18786     -0.46      0.64266
## Year2010     -0.10787      0.18685     -0.58      0.56397
## Year2011     -0.18048      0.17911     -1.01      0.31401
## Year2012     -0.20046      0.18243     -1.10      0.27228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0898, Adjusted R-squared:  0.0562
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.127  0.871  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          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 1.173 1      1.083
## LastAuthorFemale  1.213 1      1.102
## Year              1.268 16      1.007

```

Residuals from first and last author

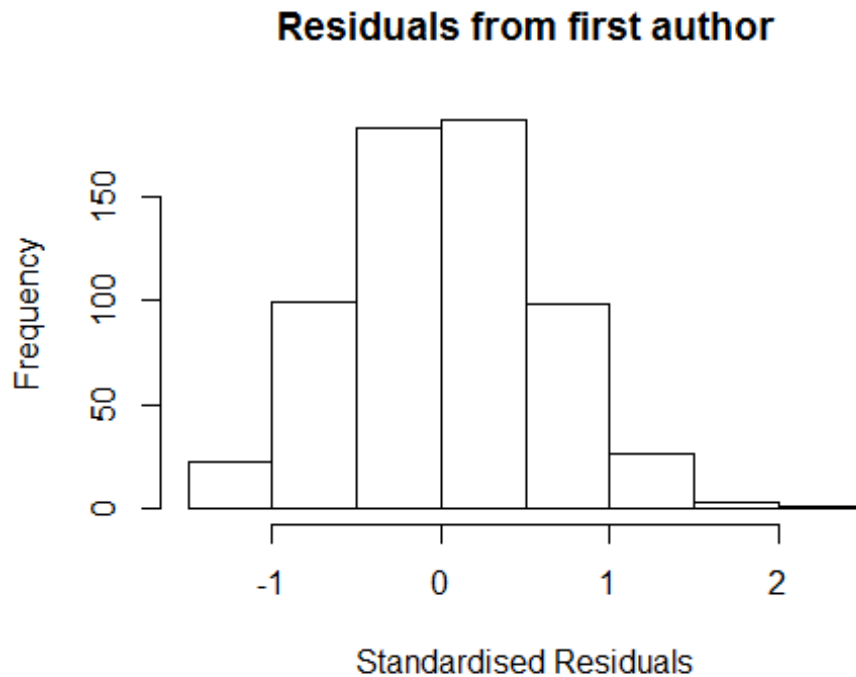


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3033 -0.3808 0.0171 0.3950 2.2387
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00386 0.15469 6.49 1.8e-10 ***
## FirstAuthorFemale1 0.10456 0.09422 1.11 0.268
## LastAuthorFemale1 0.07449 0.08052 0.92 0.355
## Year1997 0.10933 0.18220 0.60 0.549
## Year1998 -0.05685 0.19272 -0.30 0.768
## Year1999 -0.11625 0.21092 -0.55 0.582
## Year2000 0.00517 0.19769 0.03 0.979
## Year2001 0.12043 0.20001 0.60 0.547
## Year2002 0.04679 0.17750 0.26 0.792
## Year2003 -0.04912 0.17646 -0.28 0.781
## Year2004 -0.19312 0.18493 -1.04 0.297
## Year2005 -0.19799 0.17956 -1.10 0.271
```

```

## Year2006      -0.14657    0.19973   -0.73    0.463
## Year2007      -0.33594    0.18448   -1.82    0.069 .
## Year2008      -0.12753    0.17972   -0.71    0.478
## Year2009      -0.00422    0.17560   -0.02    0.981
## Year2010      -0.01541    0.17953   -0.09    0.932
## Year2011      -0.09321    0.16818   -0.55    0.580
## Year2012      -0.08132    0.16864   -0.48    0.630
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.00666
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 575 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0992 0.8820 0.9480 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      1.62e-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 1.092 1      1.045
## Year      1.092 16      1.003

```



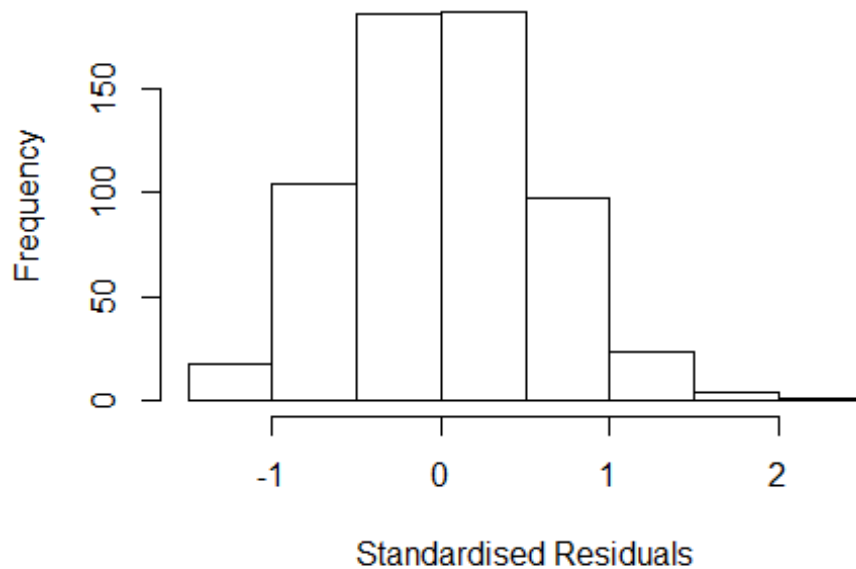
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2540 -0.3850 0.0156 0.3954 2.2250
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.005755 0.155711 6.46 2.2e-10 ***
## FirstAuthorFemale1 0.131025 0.090667 1.45 0.149
## Year1997 0.109212 0.183504 0.60 0.552
## Year1998 -0.053545 0.193097 -0.28 0.782
## Year1999 -0.110904 0.211460 -0.52 0.600
## Year2000 0.003738 0.198337 0.02 0.985
## Year2001 0.117226 0.200095 0.59 0.558
## Year2002 0.046275 0.178086 0.26 0.795
## Year2003 -0.042084 0.177897 -0.24 0.813
## Year2004 -0.193411 0.186298 -1.04 0.300
## Year2005 -0.197183 0.180872 -1.09 0.276
## Year2006 -0.136589 0.199821 -0.68 0.495
```

```

## Year2007          -0.326329    0.185420    -1.76    0.079 .
## Year2008          -0.115802    0.180167    -0.64    0.521
## Year2009           0.000128    0.176111     0.00    0.999
## Year2010          -0.001920    0.178760    -0.01    0.991
## Year2011          -0.088445    0.169359    -0.52    0.602
## Year2012          -0.077832    0.169578    -0.46    0.646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.00661
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 577 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.105  0.880  0.947  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      1.62e-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 1.13 1          1.063
## Year              1.13 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
## -1.23544 -0.37329 0.00669 0.39282 2.23693
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0060 0.1562 6.44 2.5e-10 ***
## LastAuthorFemale1 0.1080 0.0779 1.39 0.166
## Year1997 0.1206 0.1818 0.66 0.507
## Year1998 -0.0487 0.1933 -0.25 0.801
## Year1999 -0.1069 0.2113 -0.51 0.613
## Year2000 0.0128 0.1988 0.06 0.949
## Year2001 0.1215 0.2025 0.60 0.549
## Year2002 0.0589 0.1792 0.33 0.742
## Year2003 -0.0517 0.1778 -0.29 0.771
## Year2004 -0.1871 0.1865 -1.00 0.316
## Year2005 -0.1886 0.1809 -1.04 0.298
## Year2006 -0.1451 0.2004 -0.72 0.469
```

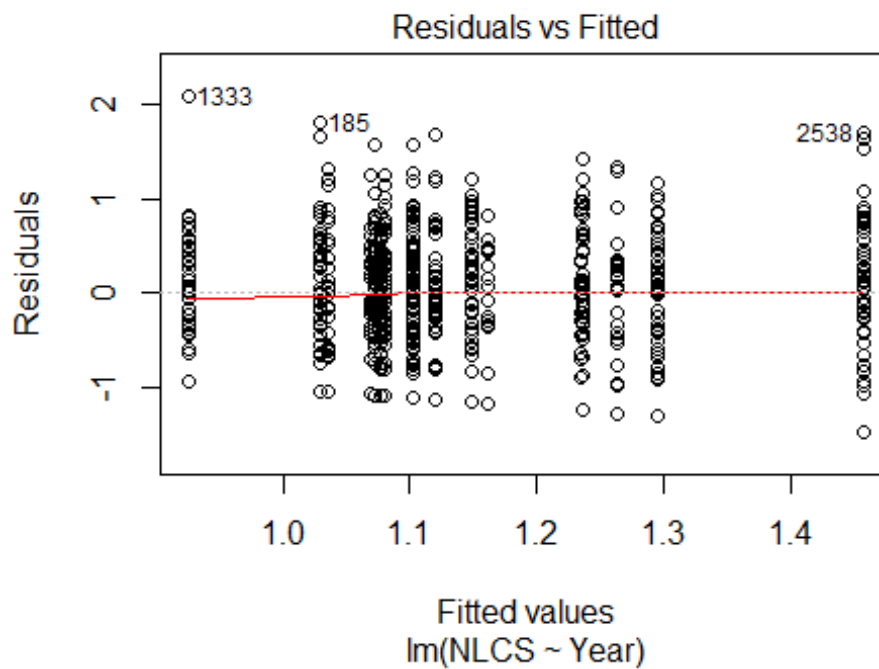
```

## Year2007          -0.3275      0.1857    -1.76     0.078 .
## Year2008          -0.1279      0.1812    -0.71     0.481
## Year2009          -0.0060      0.1777    -0.03     0.973
## Year2010          -0.0180      0.1812    -0.10     0.921
## Year2011          -0.0825      0.1694    -0.49     0.626
## Year2012          -0.0732      0.1699    -0.43     0.667
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0325, Adjusted R-squared:  0.00513
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 575 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0983 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      1.62e-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: 619"
## [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
## 136 162 114 99 109 135 110 123 107 111 119 122 127 128 115
## 2011 2012
## 121 140
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 44 15 31 25 18 42 48 42 54 57 44 41 45 51
## 2011 2012

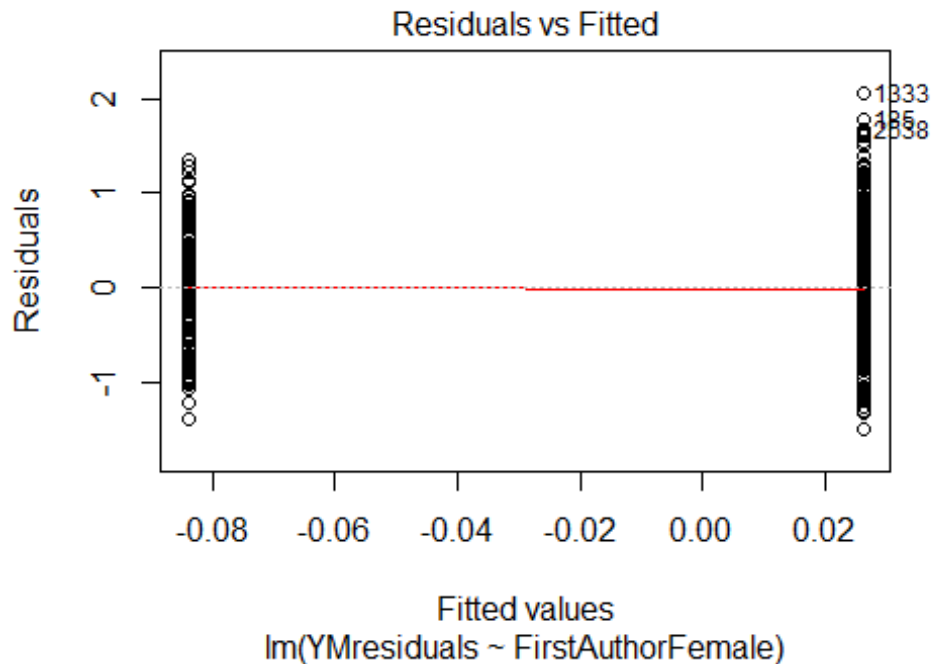
```



```
## 54 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 42 13 31 25 17 38 45 38 37 39 37 34 39 50
## 2011 2012
## 52 48
## [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 = 0.4
```

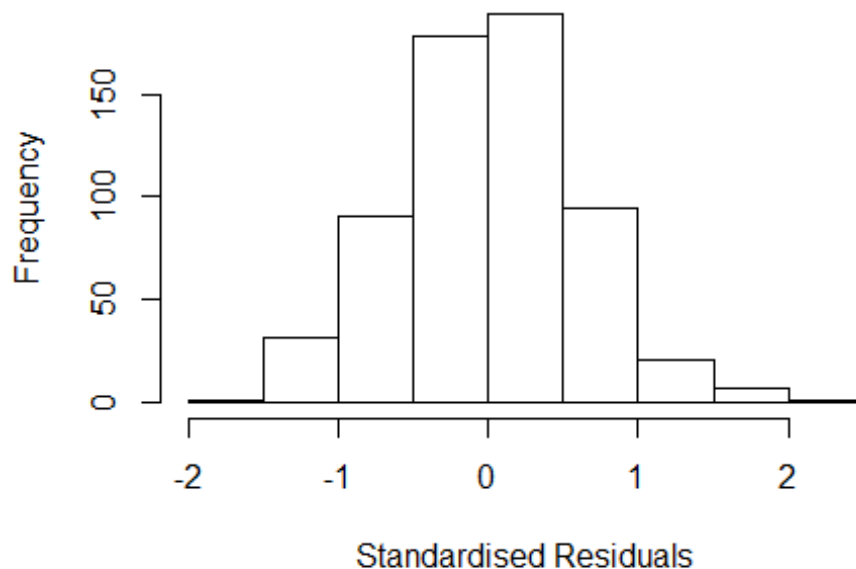


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.32, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 3.19758219305531"
## [1] "Male first author team size 2018 geometric mean: 2.87091788870654"
##
## 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] "Female last author team size 2018 geometric mean: 3.21995892709028"
## [1] "Male last author team size 2018 geometric mean: 2.87109569374073"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, 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 1.360 1      1.166
## LastAuthorFemale  1.442 1      1.201
## UniqueAuthors    1.827 4      1.078
## Year              2.132 16     1.024
```

Residuals from first and last author and team size



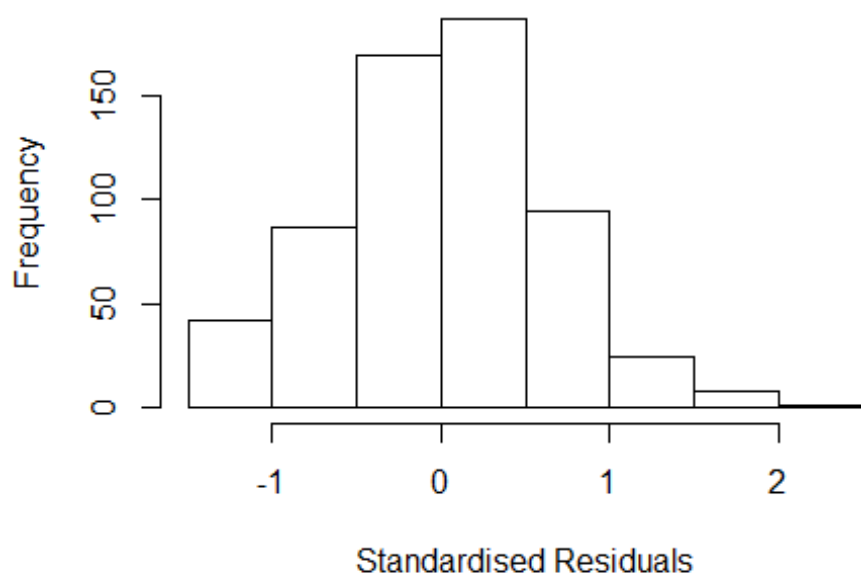
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.568 -0.391  0.016  0.392  2.021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9288    0.1468   6.33 4.9e-10 ***
## FirstAuthorFemale1 -0.1195    0.0632  -1.89 0.05925 .
## LastAuthorFemale1  0.0698    0.0763   0.91 0.36092
## UniqueAuthors2    0.2639    0.0684   3.86 0.00013 ***
## UniqueAuthors3    0.3626    0.0769   4.71 3.0e-06 ***
## UniqueAuthors4    0.1981    0.1002   1.98 0.04857 *
## UniqueAuthors5    0.3570    0.0983   3.63 0.00030 ***
## Year1997        -0.0602    0.1807  -0.33 0.73922
## Year1998         0.0810    0.2025   0.40 0.68917
## Year1999         0.0132    0.1806   0.07 0.94170
```

```

## Year2000          0.0770      0.2058      0.37  0.70855
## Year2001          0.2014      0.2107      0.96  0.33944
## Year2002          0.0293      0.1820      0.16  0.87206
## Year2003         -0.0114      0.1824     -0.06  0.95020
## Year2004         -0.2084      0.1734     -1.20  0.22977
## Year2005          0.0420      0.1734      0.24  0.80870
## Year2006         -0.0188      0.1643     -0.11  0.90892
## Year2007         -0.0422      0.1868     -0.23  0.82146
## Year2008          0.0430      0.1836      0.23  0.81515
## Year2009          0.0718      0.1845      0.39  0.69711
## Year2010          0.0946      0.1748      0.54  0.58875
## Year2011          0.1674      0.1735      0.97  0.33491
## Year2012          0.3058      0.1805      1.69  0.09067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0989, Adjusted R-squared:  0.0653
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 553 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.240  0.860  0.951  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      1.63e-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 1.318 1      1.148
## LastAuthorFemale  1.324 1      1.150
## Year              1.245 16      1.007

```

Residuals from first and last author



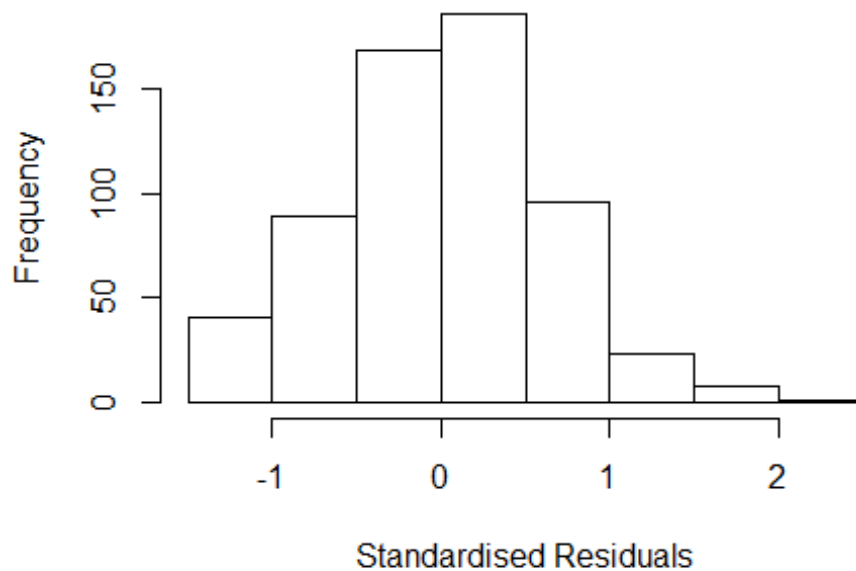
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.467 -0.394 0.025 0.400 2.114
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03846 0.15685 6.62 8e-11 ***
## FirstAuthorFemale1 -0.07977 0.06498 -1.23 0.220
## LastAuthorFemale1 0.00924 0.07671 0.12 0.904
## Year1997 -0.05622 0.18733 -0.30 0.764
## Year1998 0.14214 0.22231 0.64 0.523
## Year1999 0.05111 0.19005 0.27 0.788
## Year2000 0.23219 0.19914 1.17 0.244
## Year2001 0.21458 0.21332 1.01 0.315
## Year2002 0.07796 0.18855 0.41 0.679
## Year2003 0.05661 0.18219 0.31 0.756
## Year2004 -0.14757 0.18576 -0.79 0.427
## Year2005 0.05316 0.18098 0.29 0.769
```

```

## Year2006          0.03441    0.17301    0.20    0.842
## Year2007          0.03889    0.18920    0.21    0.837
## Year2008          0.11420    0.18876    0.60    0.545
## Year2009          0.17182    0.19467    0.88    0.378
## Year2010          0.13952    0.18731    0.74    0.457
## Year2011          0.27010    0.17956    1.50    0.133
## Year2012          0.41911    0.18708    2.24    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.0475, Adjusted R-squared:  0.0186
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 556 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.225  0.862  0.954  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      1.63e-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 1.118 1      1.057
## Year              1.118 16      1.003

```

Residuals from first author



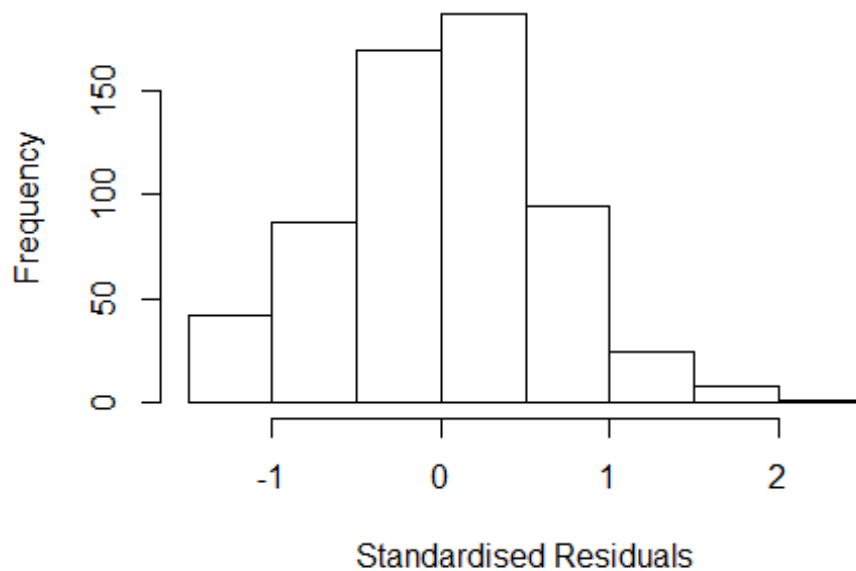
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4583 -0.3947 0.0259 0.3984 2.1128
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0399 0.1559 6.67 5.9e-11 ***
## FirstAuthorFemale1 -0.0762 0.0600 -1.27 0.204
## Year1997 -0.0570 0.1869 -0.30 0.761
## Year1998 0.1419 0.2225 0.64 0.524
## Year1999 0.0503 0.1897 0.27 0.791
## Year2000 0.2321 0.1993 1.16 0.245
## Year2001 0.2128 0.2125 1.00 0.317
## Year2002 0.0773 0.1884 0.41 0.682
## Year2003 0.0549 0.1810 0.30 0.762
## Year2004 -0.1477 0.1858 -0.79 0.427
## Year2005 0.0531 0.1810 0.29 0.770
## Year2006 0.0335 0.1726 0.19 0.846
```

```

## Year2007          0.0389      0.1893      0.21      0.837
## Year2008          0.1132      0.1881      0.60      0.547
## Year2009          0.1723      0.1946      0.89      0.376
## Year2010          0.1394      0.1873      0.74      0.457
## Year2011          0.2689      0.1788      1.50      0.133
## Year2012          0.4184      0.1867      2.24      0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.0473, Adjusted R-squared:  0.0201
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 556 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.229  0.863  0.954  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      1.63e-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 1.12 1      1.058
## Year              1.12 16      1.004

```


Residuals from last author



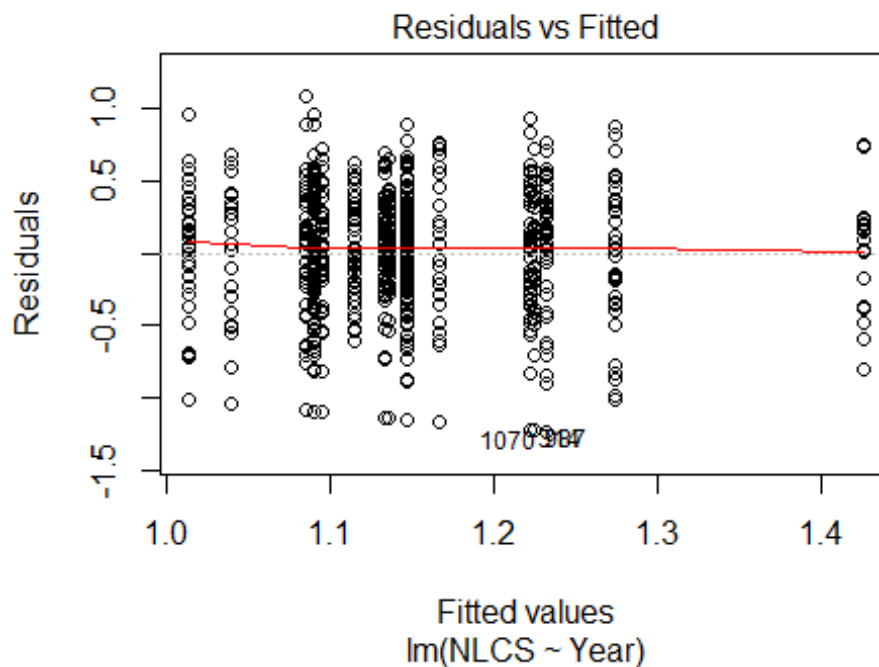
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4482 -0.3947 0.0309 0.3859 2.1259
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0301 0.1572 6.55 1.2e-10 ***
## LastAuthorFemale1 -0.0305 0.0717 -0.43 0.670
## Year1997 -0.0615 0.1871 -0.33 0.742
## Year1998 0.1515 0.2217 0.68 0.495
## Year1999 0.0497 0.1898 0.26 0.793
## Year2000 0.2303 0.2017 1.14 0.254
## Year2001 0.1905 0.2110 0.90 0.367
## Year2002 0.0735 0.1878 0.39 0.696
## Year2003 0.0523 0.1820 0.29 0.774
## Year2004 -0.1509 0.1864 -0.81 0.419
## Year2005 0.0481 0.1817 0.26 0.791
## Year2006 0.0285 0.1731 0.16 0.869
```

```

## Year2007          0.0395      0.1893      0.21      0.835
## Year2008          0.1089      0.1882      0.58      0.563
## Year2009          0.1625      0.1941      0.84      0.403
## Year2010          0.1393      0.1880      0.74      0.459
## Year2011          0.2586      0.1796      1.44      0.150
## Year2012          0.4181      0.1876      2.23      0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.621
## Multiple R-squared:  0.0451, Adjusted R-squared:  0.0178
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.218  0.859  0.958  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      1.63e-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: 612"
## [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
## 119 113 168 115 115 160 143 145 129 110 167 159 156 129 118
## 2011 2012
## 124 171
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 39 26 46 25 20 32 43 50 44 54 85 57 55 38 40
## 2011 2012

```

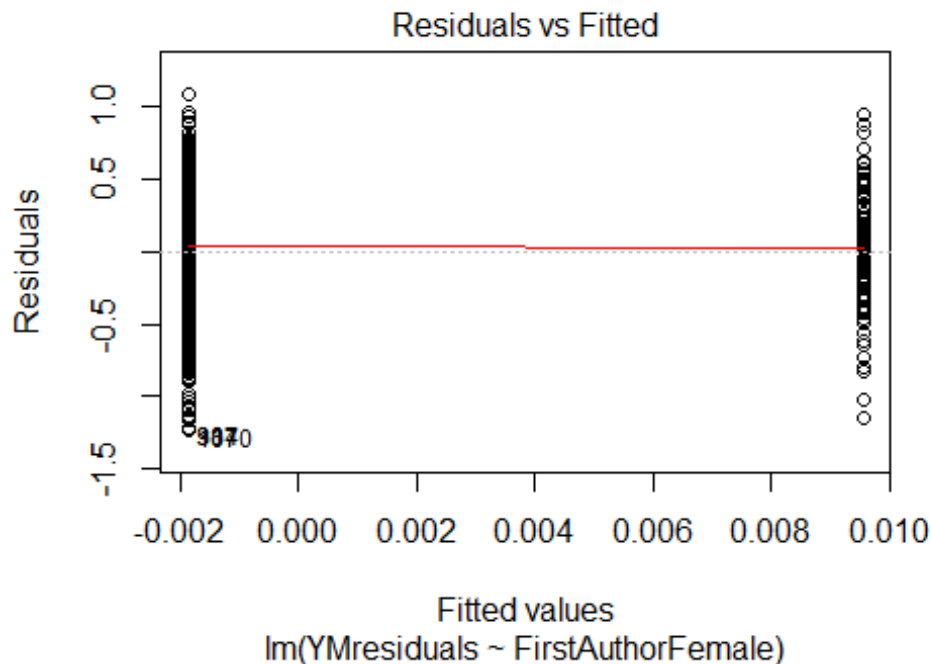
```
## 39 45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 22 41 23 16 30 40 47 39 37 63 48 44 34 35
## 2011 2012
## 36 38
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: 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 = 1.1, df = 1, p-value = 0.3
## [1] "Female first author team size 2018 geometric mean: 3.79631726035916"
## [1] "Male first author team size 2018 geometric mean: 3.04856388761912"
## 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 = 380, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.7374762132844"
## [1] "Male last author team size 2018 geometric mean: 3.37665843027842"

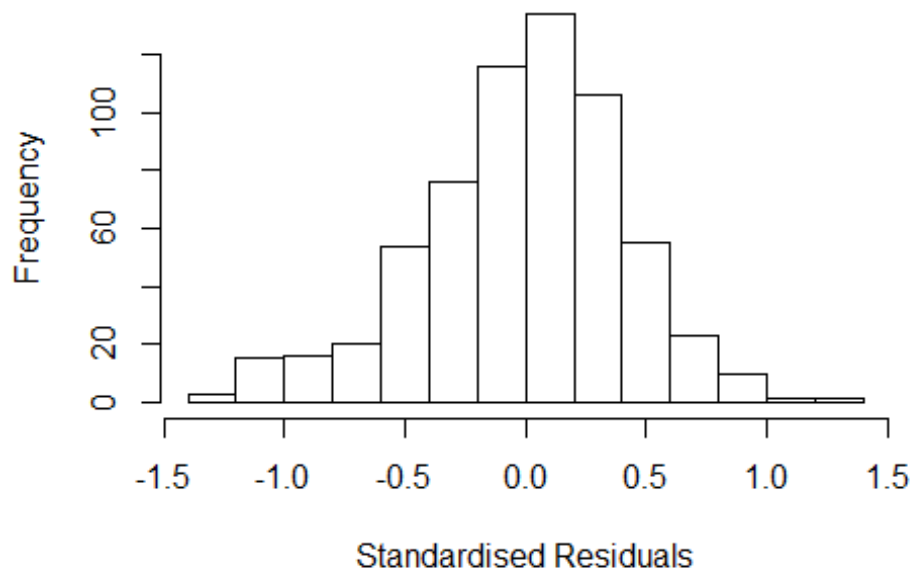
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.2
## 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	1.501	1	1.225
LastAuthorFemale	1.474	1	1.214
UniqueAuthors	1.637	4	1.064
Year	2.166	16	1.024

Residuals from first and last author and team size



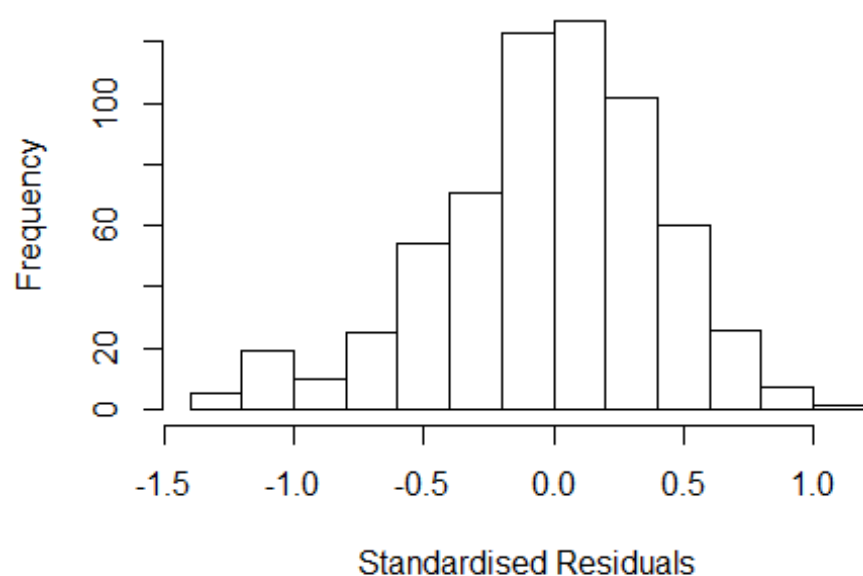
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3388 -0.2649 0.0177 0.2595 1.2116
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2159 0.1003 12.12 < 2e-16 ***
## FirstAuthorFemale1 -0.0535 0.0461 -1.16 0.24566
## LastAuthorFemale1 0.0926 0.0483 1.92 0.05548 .
## UniqueAuthors2 0.0769 0.0471 1.63 0.10287
## UniqueAuthors3 0.1658 0.0482 3.44 0.00063 ***
## UniqueAuthors4 0.1829 0.0555 3.29 0.00104 **
## UniqueAuthors5 0.1727 0.0636 2.71 0.00682 **
## Year1997 -0.1144 0.1471 -0.78 0.43696
## Year1998 -0.0787 0.1190 -0.66 0.50881
## Year1999 -0.2483 0.1470 -1.69 0.09177 .
```

```

## Year2000          0.0638      0.1421      0.45  0.65349
## Year2001         -0.0507      0.1164     -0.44  0.66335
## Year2002         -0.0236      0.1219     -0.19  0.84624
## Year2003         -0.1698      0.1076     -1.58  0.11518
## Year2004         -0.1800      0.1173     -1.53  0.12542
## Year2005         -0.2592      0.1195     -2.17  0.03051 *
## Year2006         -0.1625      0.1078     -1.51  0.13241
## Year2007         -0.1354      0.1144     -1.18  0.23713
## Year2008         -0.1114      0.1043     -1.07  0.28604
## Year2009         -0.1703      0.1288     -1.32  0.18671
## Year2010         -0.1779      0.1201     -1.48  0.13902
## Year2011         -0.2564      0.1245     -2.06  0.03978 *
## Year2012         -0.2938      0.1113     -2.64  0.00851 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0716, Adjusted R-squared:  0.038
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 566 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.859  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      1.59e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.496 1      1.223
## LastAuthorFemale  1.492 1      1.222
## Year              1.368 16      1.010

```

Residuals from first and last author



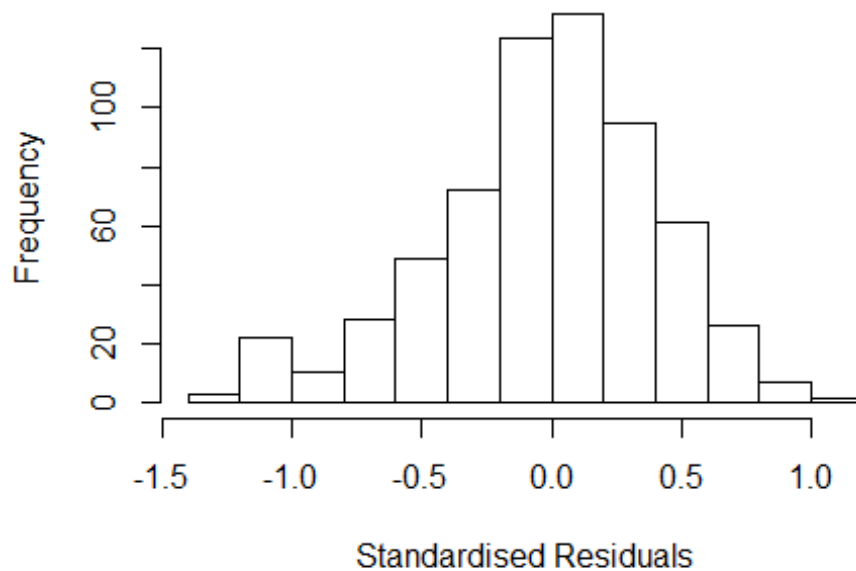
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3431 -0.2753  0.0112  0.2616  1.1044
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3012     0.0914   14.23  <2e-16 ***
## FirstAuthorFemale1 -0.0336     0.0463   -0.73    0.468
## LastAuthorFemale1  0.0924     0.0472    1.96    0.051 .
## Year1997          -0.1469     0.1519   -0.97    0.334
## Year1998          -0.0919     0.1185   -0.78    0.438
## Year1999          -0.2581     0.1429   -1.81    0.071 .
## Year2000           0.0903     0.1385    0.65    0.515
## Year2001          -0.0506     0.1120   -0.45    0.652
## Year2002          -0.0178     0.1209   -0.15    0.883
## Year2003          -0.1826     0.1034   -1.77    0.078 .
## Year2004          -0.1885     0.1157   -1.63    0.104
## Year2005          -0.2520     0.1186   -2.12    0.034 *
```

```

## Year2006          -0.1569      0.1038   -1.51    0.131
## Year2007          -0.1308      0.1107   -1.18    0.238
## Year2008          -0.1074      0.1002   -1.07    0.284
## Year2009          -0.1659      0.1250   -1.33    0.185
## Year2010          -0.1767      0.1157   -1.53    0.127
## Year2011          -0.2346      0.1202   -1.95    0.051 .
## Year2012          -0.2742      0.1089   -2.52    0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0163
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 578 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.228  0.857  0.949  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      1.59e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.179 1          1.086
## Year              1.179 16          1.005

```


Residuals from first author



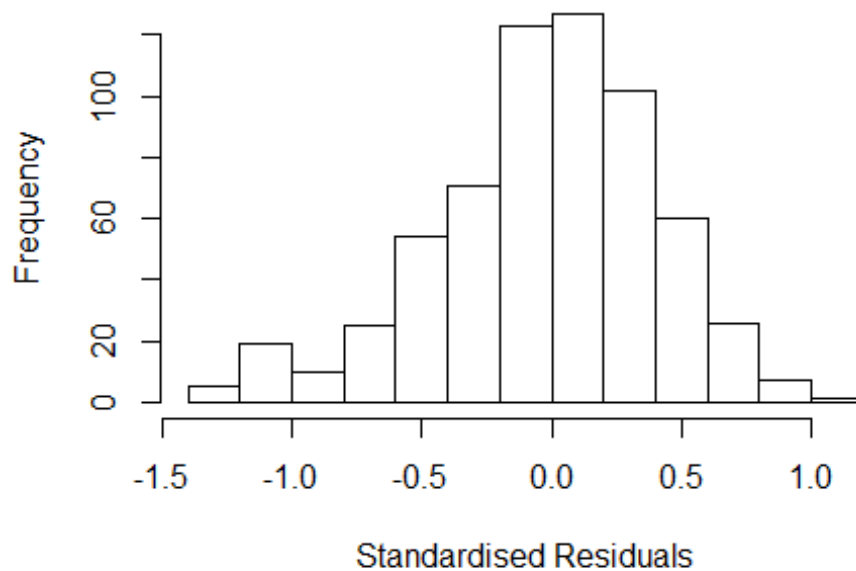
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2939 -0.2665 0.0117 0.2655 1.0984
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30776 0.09131 14.32 <2e-16 ***
## FirstAuthorFemale1 -0.00811 0.04211 -0.19 0.847
## Year1997 -0.14942 0.15037 -0.99 0.321
## Year1998 -0.08637 0.11838 -0.73 0.466
## Year1999 -0.25640 0.14241 -1.80 0.072 .
## Year2000 0.08726 0.14118 0.62 0.537
## Year2001 -0.05923 0.11131 -0.53 0.595
## Year2002 -0.01387 0.12058 -0.12 0.908
## Year2003 -0.18567 0.10406 -1.78 0.075 .
## Year2004 -0.19330 0.11502 -1.68 0.093 .
## Year2005 -0.25386 0.11909 -2.13 0.033 *
## Year2006 -0.15517 0.10346 -1.50 0.134
```

```

## Year2007          -0.14150      0.10961    -1.29      0.197
## Year2008          -0.10933      0.10020    -1.09      0.276
## Year2009          -0.16423      0.12467    -1.32      0.188
## Year2010          -0.18091      0.11594    -1.56      0.119
## Year2011          -0.23518      0.12097    -1.94      0.052 .
## Year2012          -0.26597      0.10948    -2.43      0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.0403, Adjusted R-squared:  0.0136
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 582 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.268  0.861  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      1.59e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.159 1          1.077
## Year              1.159 16          1.005

```

Residuals from last author



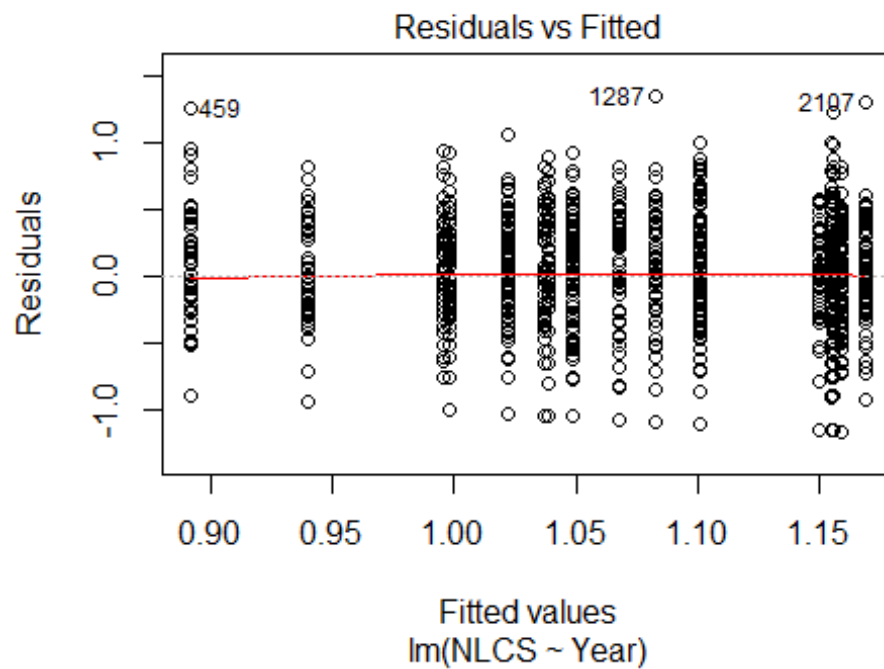
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.326 -0.282  0.012  0.265  1.108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3009     0.0913   14.24 <2e-16 ***
## LastAuthorFemale1  0.0786     0.0421    1.87  0.062 .
## Year1997        -0.1479     0.1522   -0.97  0.332
## Year1998        -0.0932     0.1182   -0.79  0.431
## Year1999        -0.2633     0.1413   -1.86  0.063 .
## Year2000         0.0867     0.1406    0.62  0.538
## Year2001        -0.0536     0.1122   -0.48  0.633
## Year2002        -0.0221     0.1201   -0.18  0.854
## Year2003        -0.1900     0.1030   -1.85  0.065 .
## Year2004        -0.1932     0.1150   -1.68  0.094 .
## Year2005        -0.2530     0.1188   -2.13  0.034 *
## Year2006        -0.1605     0.1035   -1.55  0.122
```

```

## Year2007          -0.1380      0.1096    -1.26      0.208
## Year2008          -0.1128      0.0998    -1.13      0.259
## Year2009          -0.1755      0.1239    -1.42      0.157
## Year2010          -0.1847      0.1160    -1.59      0.112
## Year2011          -0.2376      0.1201    -1.98      0.048 *
## Year2012          -0.2758      0.1092    -2.53      0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0434, Adjusted R-squared:  0.0168
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 578 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.242  0.859  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      1.59e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 630"
## [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
## 114 131 122 107 101 124 123 108 135 129 118 123 145 152 147
## 2011 2012
## 172 185
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 55 53 55 46 38 54 49 59 69 59 61 85 89 83
## 2011 2012

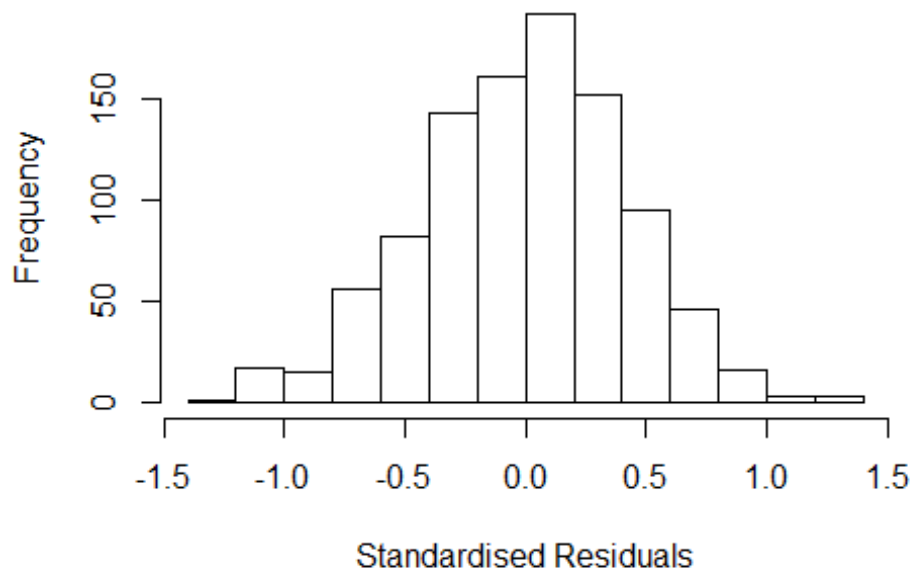
```

```
## 89 106
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 51 50 52 40 32 50 43 51 63 51 53 69 82 76
## 2011 2012
## 82 94
## [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 = 6.7, df = 1, p-value = 0.01
```


Residuals from first and last author and team size



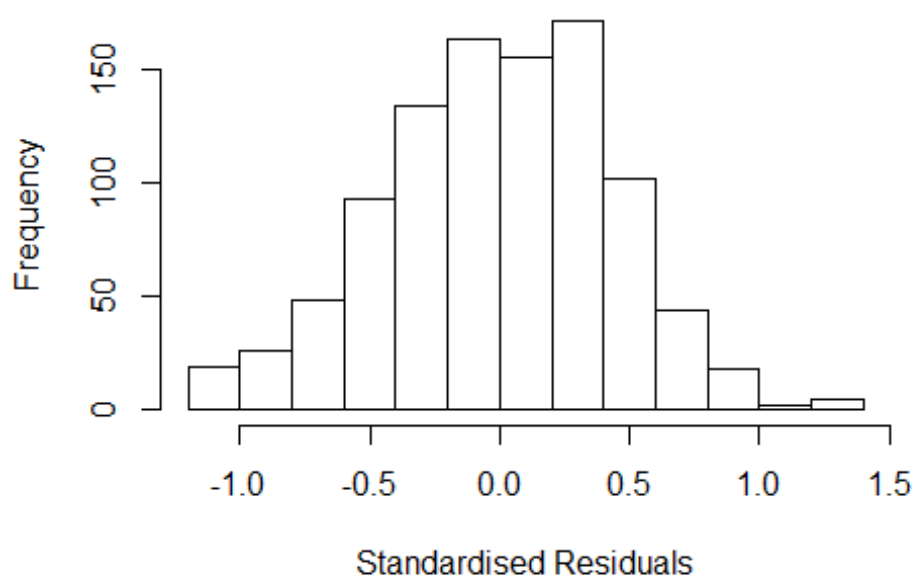
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2649 -0.2933  0.0197  0.2904  1.3753
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0574    0.0945   11.19 < 2e-16 ***
## FirstAuthorFemale1 -0.0416    0.0315   -1.32  0.18635
## LastAuthorFemale1  0.0503    0.0411    1.23  0.22080
## UniqueAuthors2     0.1523    0.0394    3.87  0.00012 ***
## UniqueAuthors3     0.2345    0.0435    5.39  9.0e-08 ***
## UniqueAuthors4     0.2217    0.0547    4.05  5.4e-05 ***
## UniqueAuthors5     0.3547    0.0566    6.26  5.6e-10 ***
## Year1997         -0.0336    0.1158   -0.29  0.77143
## Year1998         -0.2817    0.1141   -2.47  0.01376 *
## Year1999         -0.0804    0.1147   -0.70  0.48373
```

```

## Year2000          -0.1527      0.1241    -1.23   0.21869
## Year2001          -0.1257      0.1108    -1.13   0.25675
## Year2002          -0.2108      0.1096    -1.92   0.05480 .
## Year2003          -0.2054      0.1074    -1.91   0.05613 .
## Year2004          -0.1001      0.1113    -0.90   0.36885
## Year2005          -0.1960      0.0968    -2.03   0.04311 *
## Year2006          -0.0732      0.0999    -0.73   0.46389
## Year2007          -0.0773      0.1044    -0.74   0.45933
## Year2008          -0.0247      0.0955    -0.26   0.79600
## Year2009          -0.0862      0.1020    -0.85   0.39787
## Year2010          -0.1726      0.1027    -1.68   0.09323 .
## Year2011          -0.1067      0.0967    -1.10   0.26991
## Year2012          -0.2286      0.0994    -2.30   0.02163 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.0907, Adjusted R-squared:  0.0698
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 903 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.279  0.868  0.951  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          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 1.160 1          1.077
## LastAuthorFemale  1.158 1          1.076
## Year              1.227 16          1.006

```


Residuals from first and last author



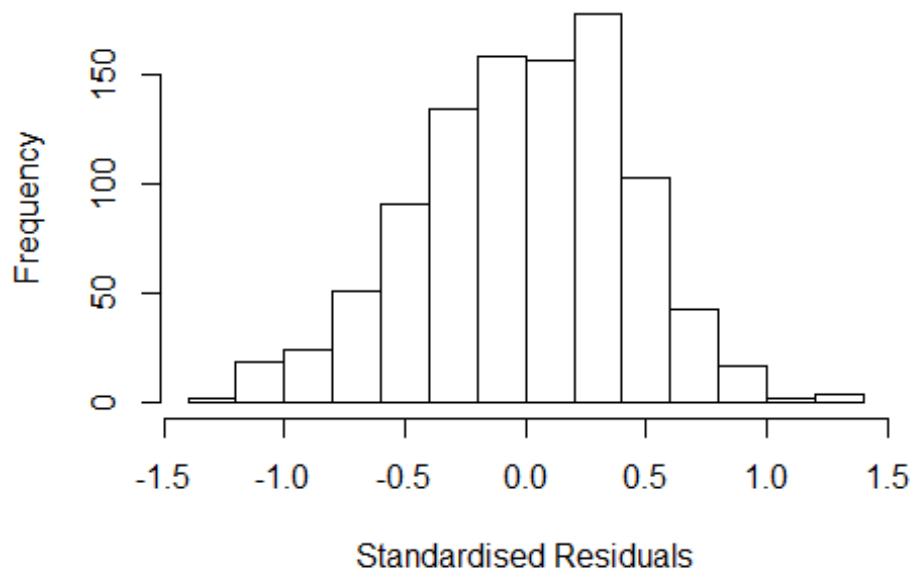
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1970 -0.3005  0.0124  0.3025  1.3431
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19701    0.07961   15.04 < 2e-16 ***
## FirstAuthorFemale1 -0.00687    0.03239   -0.21  0.83217
## LastAuthorFemale1  0.03789    0.04326    0.88  0.38130
## Year1997        -0.10055    0.11108   -0.91  0.36557
## Year1998        -0.35482    0.10592   -3.35  0.00084 ***
## Year1999        -0.10704    0.10873   -0.98  0.32514
## Year2000        -0.17756    0.12424   -1.43  0.15328
## Year2001        -0.13350    0.11075   -1.21  0.22834
## Year2002        -0.21600    0.10344   -2.09  0.03705 *
## Year2003        -0.23228    0.10267   -2.26  0.02390 *
## Year2004        -0.11114    0.10753   -1.03  0.30160
## Year2005        -0.20992    0.09059   -2.32  0.02069 *
```

```

## Year2006          -0.04642      0.09191    -0.51  0.61361
## Year2007          -0.08563      0.09937    -0.86  0.38902
## Year2008          -0.02488      0.08934    -0.28  0.78067
## Year2009          -0.05073      0.09588    -0.53  0.59687
## Year2010          -0.17878      0.09472    -1.89  0.05938 .
## Year2011          -0.04411      0.09033    -0.49  0.62543
## Year2012          -0.19481      0.09350    -2.08  0.03746 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.0393, Adjusted R-squared:  0.0213
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 893 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.336  0.872  0.949  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      1.02e-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 1.118 1          1.058
## Year              1.118 16          1.004

```

Residuals from first author



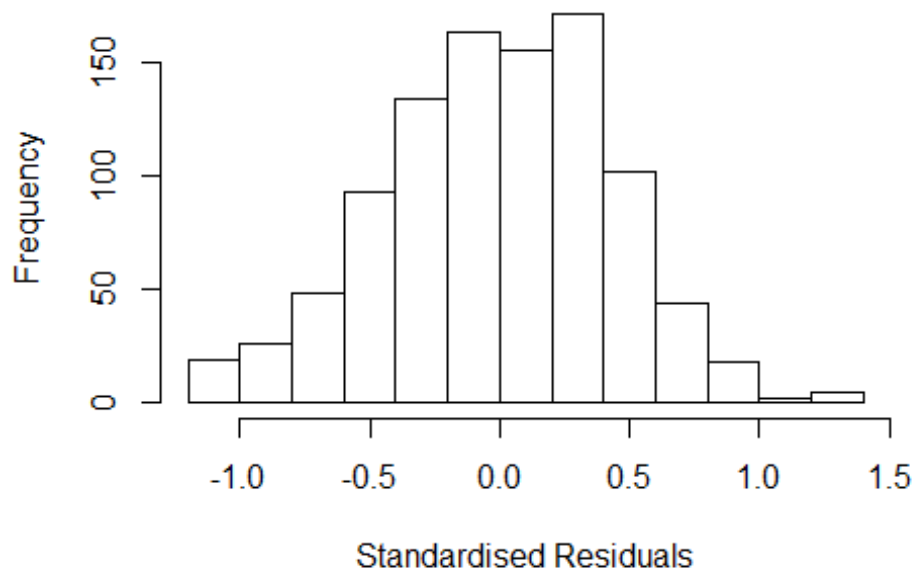
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2013 -0.3001 0.0122 0.3016 1.3411
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20126 0.07940 15.13 < 2e-16 ***
## FirstAuthorFemale1 0.00113 0.03185 0.04 0.97180
## Year1997 -0.10177 0.11131 -0.91 0.36076
## Year1998 -0.35508 0.10638 -3.34 0.00088 ***
## Year1999 -0.11095 0.10834 -1.02 0.30609
## Year2000 -0.17912 0.12424 -1.44 0.14972
## Year2001 -0.13879 0.11052 -1.26 0.20951
## Year2002 -0.21855 0.10314 -2.12 0.03435 *
## Year2003 -0.23669 0.10244 -2.31 0.02107 *
## Year2004 -0.11341 0.10735 -1.06 0.29106
## Year2005 -0.21230 0.09068 -2.34 0.01943 *
## Year2006 -0.04639 0.09213 -0.50 0.61470
```

```

## Year2007          -0.08545    0.09945   -0.86  0.39046
## Year2008          -0.02259    0.08958   -0.25  0.80097
## Year2009          -0.04917    0.09605   -0.51  0.60878
## Year2010          -0.17895    0.09482   -1.89  0.05942 .
## Year2011          -0.04618    0.09040   -0.51  0.60958
## Year2012          -0.19553    0.09375   -2.09  0.03728 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.0385, Adjusted R-squared:  0.0215
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 890 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.340  0.871  0.948  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      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 1.117 1          1.057
## Year              1.117 16          1.003

```

Residuals from last author



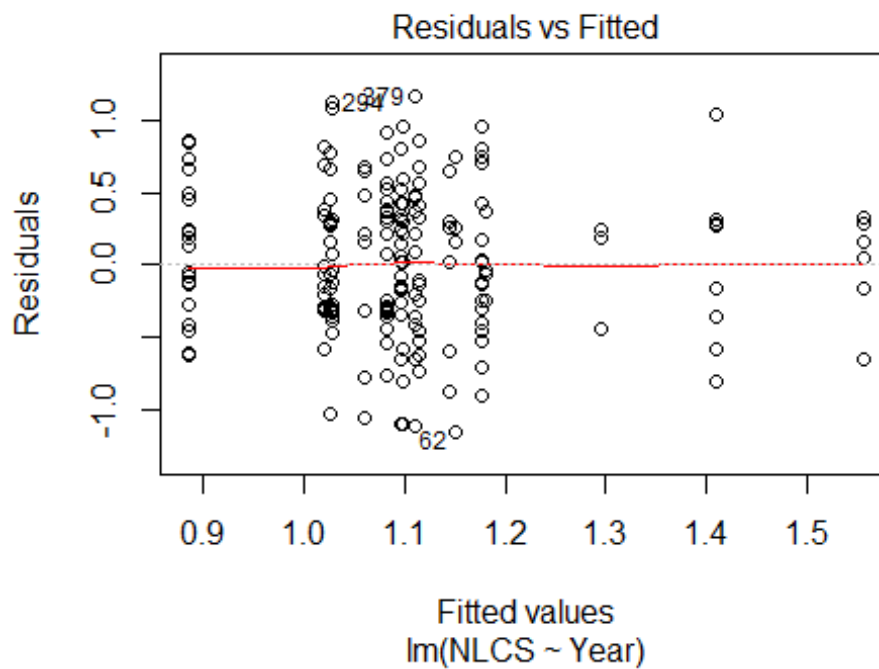
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1964 -0.2991 0.0137 0.3010 1.3452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1964 0.0795 15.05 < 2e-16 ***
## LastAuthorFemale1 0.0356 0.0426 0.84 0.40265
## Year1997 -0.1005 0.1109 -0.91 0.36503
## Year1998 -0.3544 0.1058 -3.35 0.00084 ***
## Year1999 -0.1079 0.1085 -0.99 0.32024
## Year2000 -0.1775 0.1243 -1.43 0.15365
## Year2001 -0.1340 0.1107 -1.21 0.22626
## Year2002 -0.2162 0.1034 -2.09 0.03675 *
## Year2003 -0.2330 0.1025 -2.27 0.02323 *
## Year2004 -0.1125 0.1067 -1.05 0.29197
## Year2005 -0.2108 0.0904 -2.33 0.01997 *
## Year2006 -0.0476 0.0916 -0.52 0.60337
```

```

## Year2007          -0.0865      0.0992   -0.87  0.38317
## Year2008          -0.0257      0.0892   -0.29  0.77373
## Year2009          -0.0522      0.0953   -0.55  0.58390
## Year2010          -0.1806      0.0940   -1.92  0.05501 .
## Year2011          -0.0456      0.0902   -0.51  0.61338
## Year2012          -0.1953      0.0934   -2.09  0.03686 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.0392, Adjusted R-squared:  0.0223
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 890 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.872  0.949  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      1.02e-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: 982"
## [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
##   18   14   13   12   14   19   22   16   19   25   33   22   28   29   22
## 2011 2012
##   31   40
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    3    4    4    7    8   12    8    9   15   24   13   19   20   11
## 2011 2012

```

```
## 13 25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 3 3 4 7 8 11 7 8 13 20 12 17 19 10
## 2011 2012
## 12 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 = 11, df = 16, p-value = 0.8
```



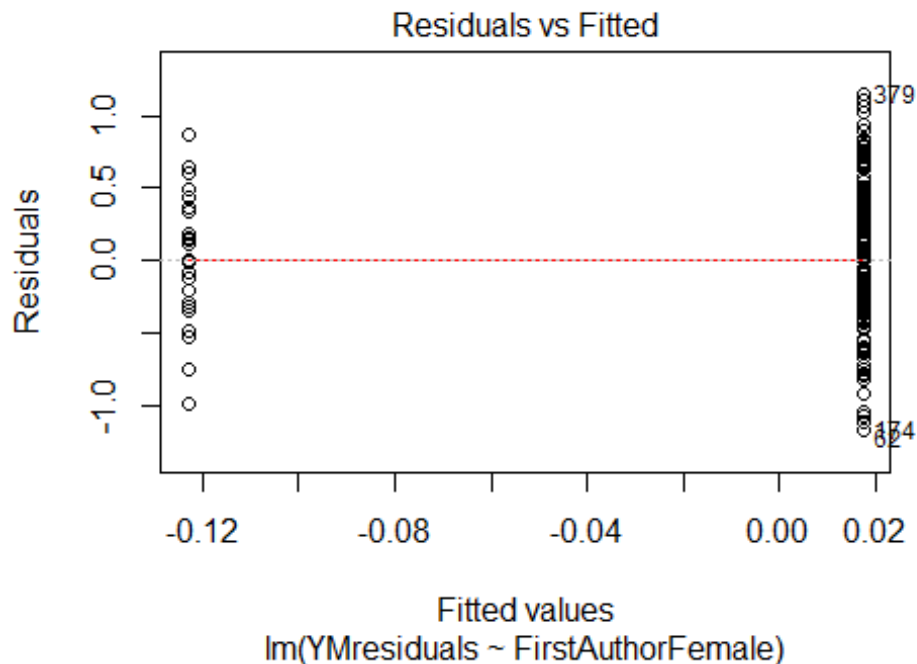
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.72, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 1.64375182951723"
## [1] "Male first author team size 2018 geometric mean: 3.25933561120894"

## 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 = 32, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.68084338963949"
## [1] "Male last author team size 2018 geometric mean: 3.21120783068661"

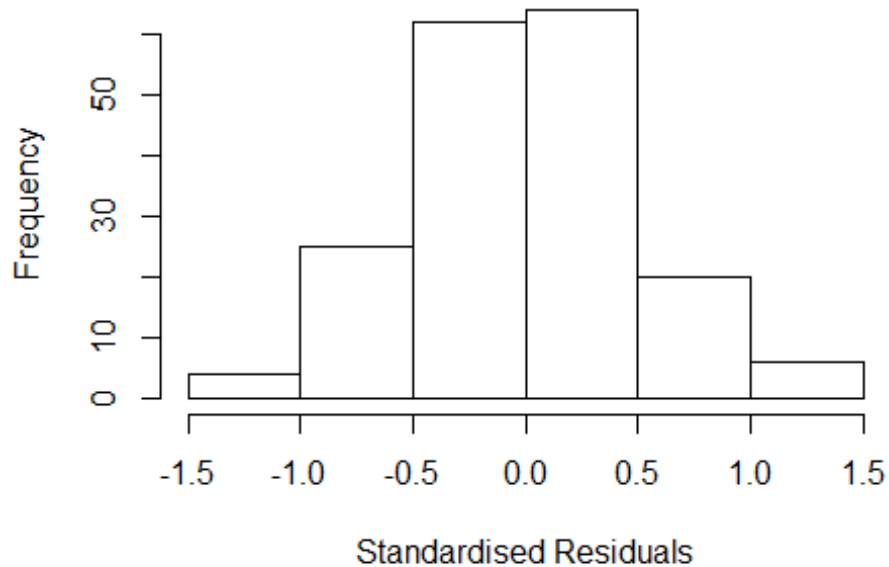
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 37, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
## FirstAuthorFemale	2.969	1	1.723
## LastAuthorFemale	2.801	1	1.674
## UniqueAuthors	6.224	4	1.257
## Year	8.626	16	1.070

Residuals from first and last author and team size



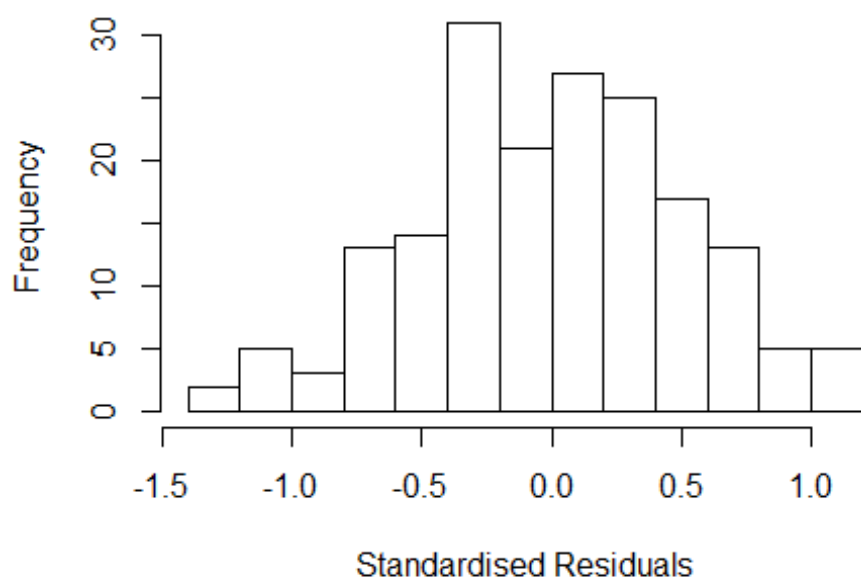
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3033 -0.3539 -0.0101 0.3456 1.3408
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.392 0.184 7.55 3.3e-12 ***
## FirstAuthorFemale1 -0.138 0.145 -0.95 0.3442
## LastAuthorFemale1 0.103 0.143 0.72 0.4720
## UniqueAuthors2 0.111 0.110 1.00 0.3172
## UniqueAuthors3 0.304 0.110 2.77 0.0063 **
## UniqueAuthors4 0.118 0.259 0.46 0.6481
## UniqueAuthors5 0.414 0.164 2.52 0.0128 *
## Year1997 -0.162 0.276 -0.59 0.5592
## Year1998 -0.311 0.242 -1.28 0.2011
## Year1999 -0.143 0.453 -0.32 0.7519
```

```

## Year2000          -0.353      0.267    -1.32    0.1878
## Year2001          -0.393      0.326    -1.21    0.2294
## Year2002          -0.427      0.244    -1.75    0.0814 .
## Year2003          -0.164      0.309    -0.53    0.5956
## Year2004          -0.513      0.318    -1.61    0.1091
## Year2005          -0.579      0.194    -2.98    0.0033 **
## Year2006          -0.617      0.211    -2.93    0.0039 **
## Year2007          -0.584      0.253    -2.31    0.0223 *
## Year2008          -0.390      0.206    -1.90    0.0597 .
## Year2009          -0.424      0.218    -1.95    0.0533 .
## Year2010          -0.280      0.247    -1.13    0.2595
## Year2011          -0.466      0.231    -2.01    0.0457 *
## Year2012          -0.518      0.212    -2.45    0.0154 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.137, Adjusted R-squared:  0.0168
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 175 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.480  0.893  0.956  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      5.52e-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 2.355 1      1.535
## LastAuthorFemale  2.491 1      1.578
## Year              1.867 16      1.020

```

Residuals from first and last author



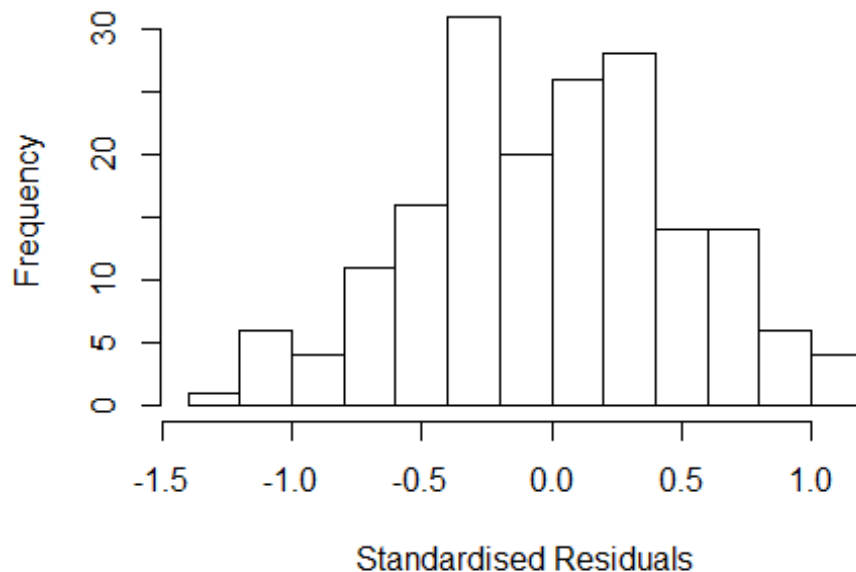
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2829 -0.3258 0.0141 0.3506 1.1668
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5013 0.1501 10.00 < 2e-16 ***
## FirstAuthorFemale1 -0.2154 0.1340 -1.61 0.10998
## LastAuthorFemale1 0.1956 0.1458 1.34 0.18164
## Year1997 -0.1988 0.2407 -0.83 0.40998
## Year1998 -0.3092 0.2144 -1.44 0.15108
## Year1999 -0.2185 0.4054 -0.54 0.59077
## Year2000 -0.3107 0.2358 -1.32 0.18952
## Year2001 -0.3879 0.2941 -1.32 0.18904
## Year2002 -0.3483 0.2375 -1.47 0.14445
## Year2003 -0.0821 0.2972 -0.28 0.78276
## Year2004 -0.4324 0.3417 -1.27 0.20752
## Year2005 -0.5960 0.1757 -3.39 0.00087 ***
```

```

## Year2006          -0.5645      0.1944    -2.90  0.00421 **
## Year2007          -0.5191      0.2179    -2.38  0.01839 *
## Year2008          -0.3527      0.1876    -1.88  0.06193 .
## Year2009          -0.3730      0.2064    -1.81  0.07251 .
## Year2010          -0.2357      0.2541    -0.93  0.35504
## Year2011          -0.4978      0.2176    -2.29  0.02349 *
## Year2012          -0.4495      0.1918    -2.34  0.02034 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0789, Adjusted R-squared:  -0.0234
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 170 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.545  0.882  0.958  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      5.52e-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 1.525 1      1.235
## Year      1.525 16      1.013

```

Residuals from first author



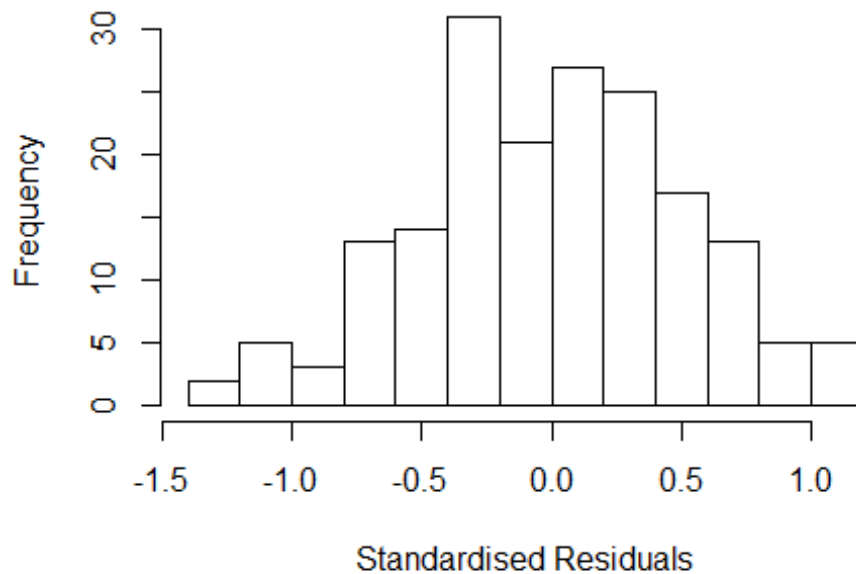
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3139 -0.3355  0.0111  0.3479  1.1579
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5017     0.1502   10.00  <2e-16 ***
## FirstAuthorFemale1 -0.1191     0.1054   -1.13   0.2602
## Year1997          -0.1989     0.2410   -0.83   0.4103
## Year1998          -0.3096     0.2146   -1.44   0.1509
## Year1999          -0.1877     0.4294   -0.44   0.6626
## Year2000          -0.3189     0.2438   -1.31   0.1927
## Year2001          -0.3860     0.2952   -1.31   0.1929
## Year2002          -0.3487     0.2374   -1.47   0.1438
## Year2003          -0.0829     0.2968   -0.28   0.7803
## Year2004          -0.3928     0.3460   -1.14   0.2580
## Year2005          -0.5585     0.1845   -3.03   0.0029 **
## Year2006          -0.5635     0.1954   -2.88   0.0045 **
```

```

## Year2007          -0.5106      0.2185   -2.34   0.0207 *
## Year2008          -0.3226      0.1903   -1.70   0.0919 .
## Year2009          -0.3528      0.2017   -1.75   0.0821 .
## Year2010          -0.2706      0.2429   -1.11   0.2669
## Year2011          -0.4969      0.2178   -2.28   0.0238 *
## Year2012          -0.4382      0.1913   -2.29   0.0232 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.527
## Multiple R-squared:  0.0703, Adjusted R-squared:  -0.0267
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 166 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.514  0.870   0.957   0.912  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      5.52e-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 1.395 1          1.181
## Year            1.395 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
## -1.269 -0.304 -0.001 0.350 1.211
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5017 0.1502 10.00 < 2e-16 ***
## LastAuthorFemale1 0.0774 0.1218 0.64 0.52589
## Year1997 -0.1989 0.2410 -0.83 0.41030
## Year1998 -0.3097 0.2146 -1.44 0.15085
## Year1999 -0.2398 0.3850 -0.62 0.53419
## Year2000 -0.3267 0.2540 -1.29 0.20016
## Year2001 -0.3858 0.2953 -1.31 0.19331
## Year2002 -0.3488 0.2384 -1.46 0.14545
## Year2003 -0.0830 0.2969 -0.28 0.78012
## Year2004 -0.4078 0.3466 -1.18 0.24102
## Year2005 -0.6001 0.1787 -3.36 0.00097 ***
## Year2006 -0.5810 0.1975 -2.94 0.00374 **
```

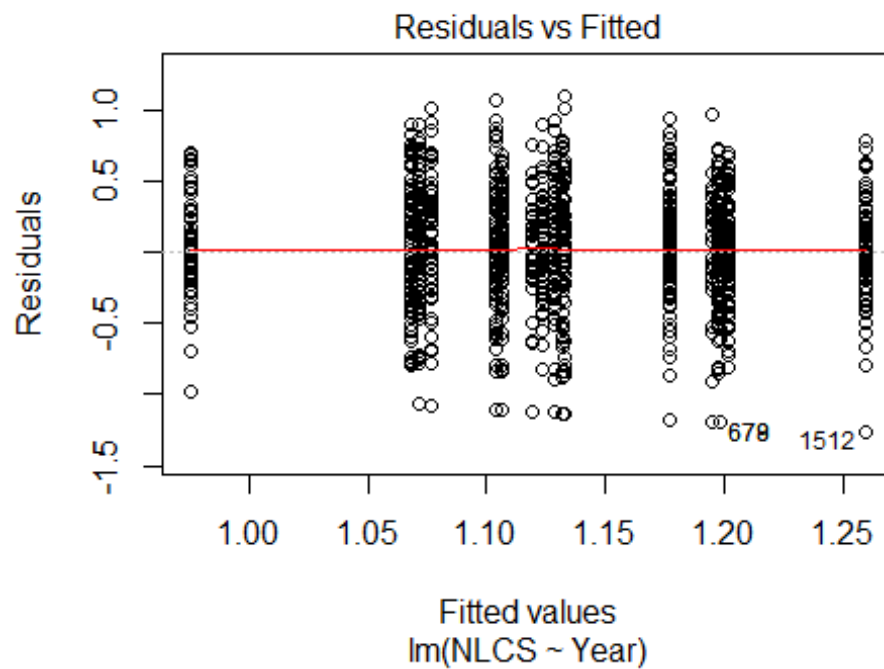
```

## Year2007          -0.5641      0.2179   -2.59  0.01050 *
## Year2008          -0.3875      0.1841   -2.10  0.03684 *
## Year2009          -0.3836      0.2050   -1.87  0.06316 .
## Year2010          -0.3097      0.2462   -1.26  0.21029
## Year2011          -0.4968      0.2178   -2.28  0.02385 *
## Year2012          -0.4728      0.1897   -2.49  0.01370 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.0683, Adjusted R-squared:  -0.0289
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.540  0.877  0.956  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      5.52e-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: 181"
## [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
##   80   95  134  127   91  149  119   98  104  120  145  181  178  134  160
## 2011 2012
##  197  189
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   40   60   58   34   54   70   69   63   74   92   92  113   88  111
## 2011 2012

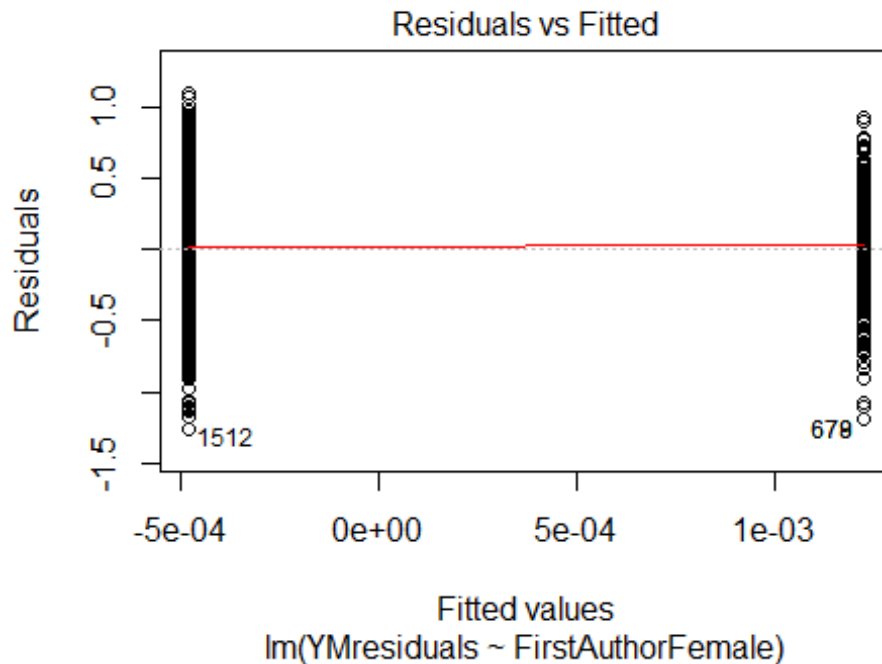
```



```
## 117 118
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 32 36 51 56 30 52 66 66 51 72 82 82 99 82 104
## 2011 2012
## 107 113
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```

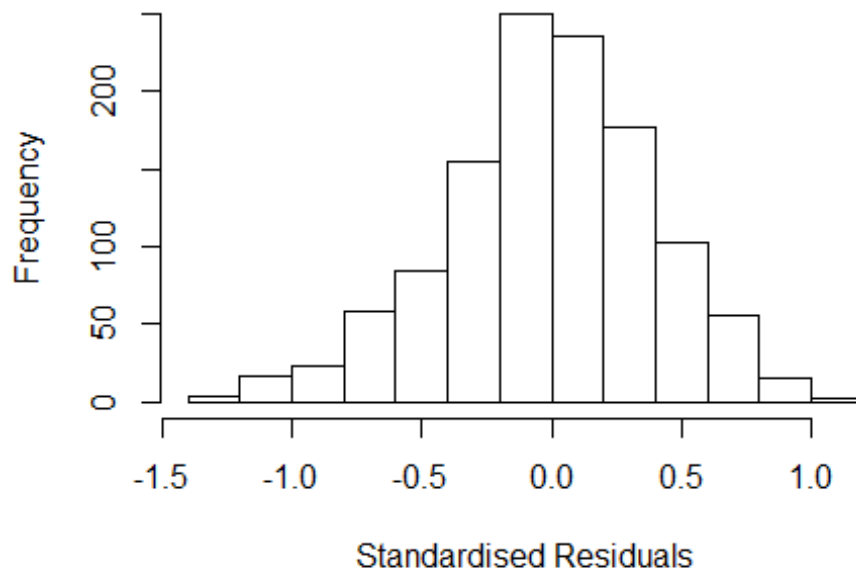


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.72, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.17817113238142"
## [1] "Male first author team size 2018 geometric mean: 2.40008285892197"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.66779042524817"
## [1] "Male last author team size 2018 geometric mean: 2.73515801508824"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 9e-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 1.240 1      1.114
## LastAuthorFemale  1.241 1      1.114
## UniqueAuthors     1.366 4      1.040
## Year              1.620 16     1.015
```

Residuals from first and last author and team size



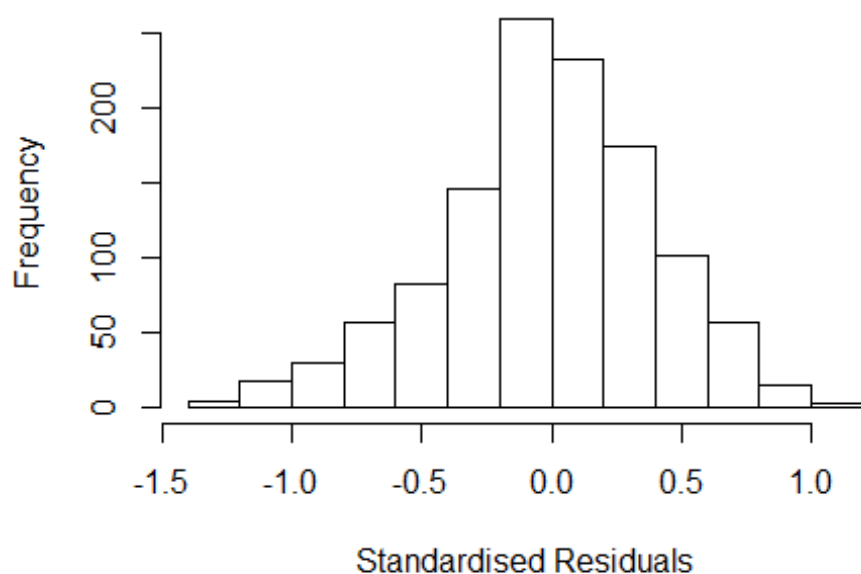
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24817 -0.24878 -0.00163 0.24556 1.12573
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10527 0.11505 9.61 < 2e-16 ***
## FirstAuthorFemale1 0.00864 0.02777 0.31 0.75582
## LastAuthorFemale1 -0.00257 0.02964 -0.09 0.93085
## UniqueAuthors2 0.07891 0.03119 2.53 0.01154 *
## UniqueAuthors3 0.13039 0.03559 3.66 0.00026 ***
## UniqueAuthors4 0.09571 0.04164 2.30 0.02172 *
## UniqueAuthors5 0.11892 0.04695 2.53 0.01144 *
## Year1997 -0.05595 0.12629 -0.44 0.65785
## Year1998 0.02217 0.12835 0.17 0.86290
## Year1999 0.01965 0.13293 0.15 0.88250
```

```

## Year2000      0.05979      0.13408      0.45  0.65572
## Year2001      0.04900      0.12138      0.40  0.68650
## Year2002     -0.07442      0.12650     -0.59  0.55644
## Year2003     -0.17806      0.12277     -1.45  0.14721
## Year2004     -0.05010      0.12432     -0.40  0.68701
## Year2005      0.03208      0.12066      0.27  0.79041
## Year2006      0.09894      0.11734      0.84  0.39931
## Year2007      0.01844      0.11844      0.16  0.87628
## Year2008      0.01251      0.11895      0.11  0.91626
## Year2009     -0.07139      0.12215     -0.58  0.55904
## Year2010     -0.00391      0.11873     -0.03  0.97372
## Year2011     -0.10918      0.11838     -0.92  0.35655
## Year2012     -0.08264      0.11880     -0.70  0.48678
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0466, Adjusted R-squared:  0.0285
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 1090 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.854  0.954  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      8.47e-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 1.213 1      1.101
## LastAuthorFemale 1.212 1      1.101
## Year      1.225 16      1.006

```

Residuals from first and last author



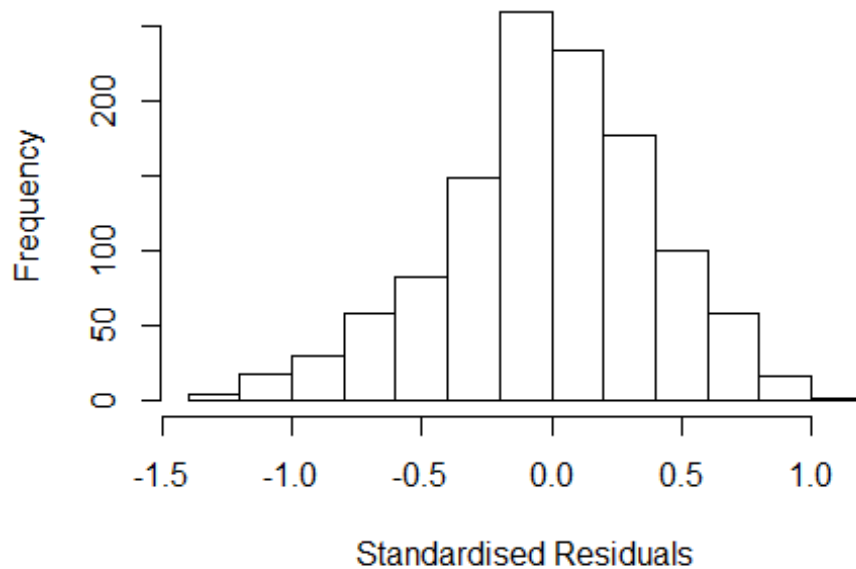
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27855 -0.24741 -0.00915  0.24236  1.05431
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.17669    0.11259   10.45  <2e-16 ***
## FirstAuthorFemale1  0.01176    0.02750    0.43    0.67
## LastAuthorFemale1 -0.01854    0.02951   -0.63    0.53
## Year1997          -0.06862    0.12588   -0.55    0.59
## Year1998          -0.00323    0.13013   -0.02    0.98
## Year1999           0.01527    0.13199    0.12    0.91
## Year2000           0.04216    0.13639    0.31    0.76
## Year2001           0.05240    0.12143    0.43    0.67
## Year2002          -0.07828    0.12731   -0.61    0.54
## Year2003          -0.19605    0.12284   -1.60    0.11
## Year2004          -0.06593    0.12457   -0.53    0.60
## Year2005           0.02930    0.12105    0.24    0.81
```

```

## Year2006      0.10186    0.11733    0.87    0.39
## Year2007      0.01654    0.11865    0.14    0.89
## Year2008      0.00673    0.11867    0.06    0.95
## Year2009     -0.07060    0.12289   -0.57    0.57
## Year2010     -0.00833    0.11863   -0.07    0.94
## Year2011     -0.10858    0.11870   -0.91    0.36
## Year2012     -0.07745    0.11852   -0.65    0.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0312, Adjusted R-squared:  0.0162
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1093 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.231  0.855   0.954   0.895   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      8.47e-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 1.106 1      1.052
## Year              1.106 16      1.003

```

Residuals from first author



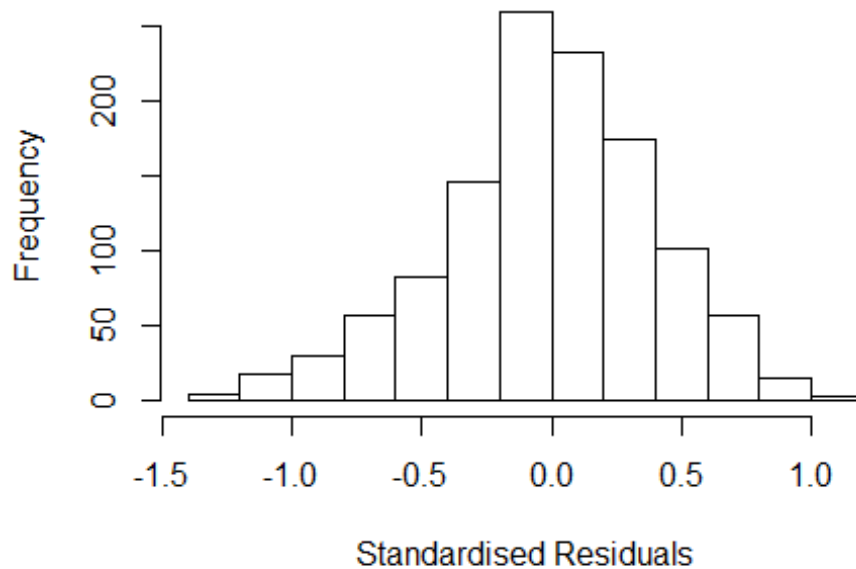
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27443 -0.24639 -0.00555  0.24156  1.05911
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.171889   0.111026  10.56  <2e-16 ***
## FirstAuthorFemale1  0.005007   0.026445   0.19   0.85
## Year1997        -0.064090   0.124606  -0.51   0.61
## Year1998        -0.000981   0.129679  -0.01   0.99
## Year1999         0.019859   0.130719   0.15   0.88
## Year2000         0.045149   0.135395   0.33   0.74
## Year2001         0.058542   0.119497   0.49   0.62
## Year2002        -0.075305   0.126288  -0.60   0.55
## Year2003        -0.191900   0.121700  -1.58   0.12
## Year2004        -0.063152   0.123694  -0.51   0.61
## Year2005         0.030768   0.120332   0.26   0.80
## Year2006         0.102543   0.116855   0.88   0.38
```

```

## Year2007          0.018725    0.117926    0.16      0.87
## Year2008          0.008500    0.117978    0.07      0.94
## Year2009         -0.068682    0.122179   -0.56      0.57
## Year2010         -0.005688    0.117716   -0.05      0.96
## Year2011         -0.106166    0.117924   -0.90      0.37
## Year2012         -0.077345    0.118162   -0.65      0.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.031, Adjusted R-squared:  0.0168
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 1086 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.233  0.853   0.953   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      8.47e-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 1.108 1          1.053
## Year            1.108 16          1.003

```


Residuals from last author



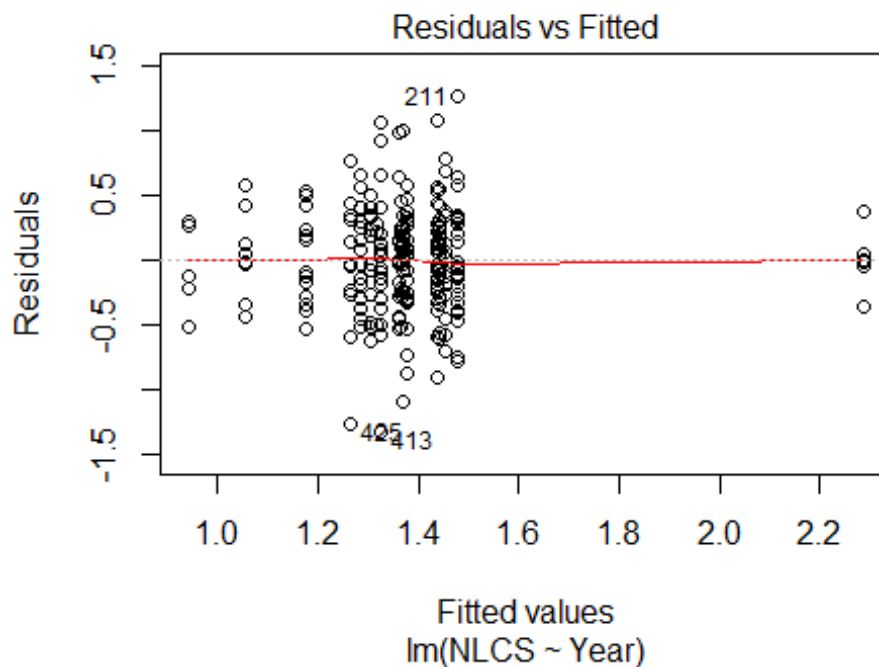
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28089 -0.24552 -0.00714 0.24451 1.05411
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17689 0.11269 10.44 <2e-16 ***
## LastAuthorFemale1 -0.01362 0.02834 -0.48 0.63
## Year1997 -0.06786 0.12586 -0.54 0.59
## Year1998 -0.00187 0.13007 -0.01 0.99
## Year1999 0.01943 0.13074 0.15 0.88
## Year2000 0.04357 0.13627 0.32 0.75
## Year2001 0.05500 0.12124 0.45 0.65
## Year2002 -0.07748 0.12729 -0.61 0.54
## Year2003 -0.19431 0.12273 -1.58 0.11
## Year2004 -0.06340 0.12409 -0.51 0.61
## Year2005 0.03025 0.12101 0.25 0.80
## Year2006 0.10400 0.11710 0.89 0.37
```

```

## Year2007      0.01950      0.11823      0.16      0.87
## Year2008      0.00929      0.11819      0.08      0.94
## Year2009     -0.06808      0.12250     -0.56      0.58
## Year2010     -0.00451      0.11800     -0.04      0.97
## Year2011     -0.10698      0.11868     -0.90      0.37
## Year2012     -0.07593      0.11836     -0.64      0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0311, Adjusted R-squared:  0.0169
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1089 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.228  0.854  0.952  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      8.47e-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: 1181"
## [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
##   22  16  42  17  14  38  28   5  19  28  23  28  24  18  33
## 2011 2012
##   15  31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12   6  25   9   6  12  23   3  17  18  14  20  18  12  14
## 2011 2012

```

```
##      7    22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   11    5   23    7    6    9   20    3   16   11   13   19   17   12   12
## 2011 2012
##    7   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 = 13, df = 16, p-value = 0.6
```



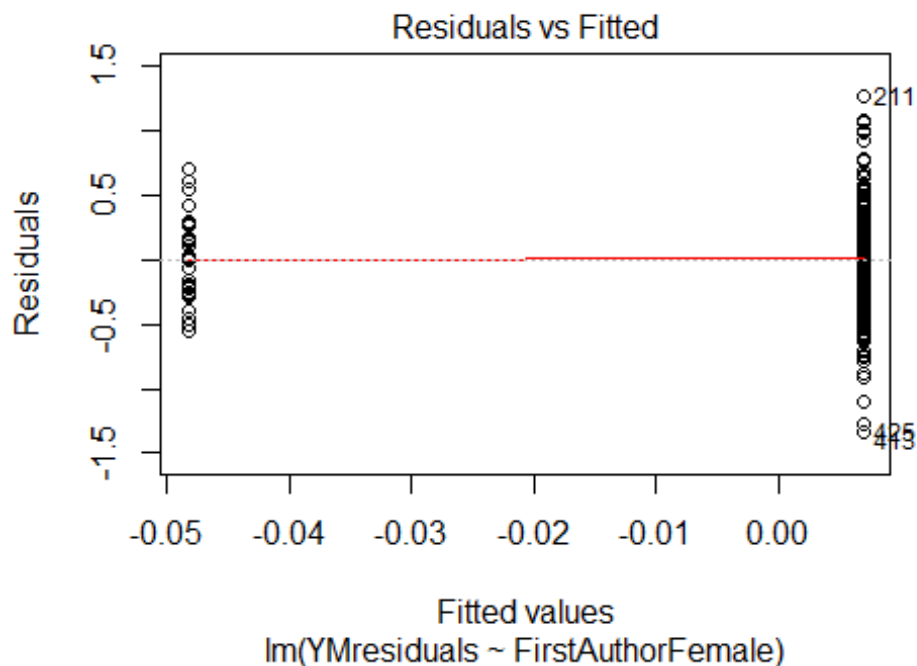
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 3.01193170922985"
## [1] "Male first author team size 2018 geometric mean: 3.54956446195497"

## 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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.29311609465006"
## [1] "Male last author team size 2018 geometric mean: 3.3468152310048"

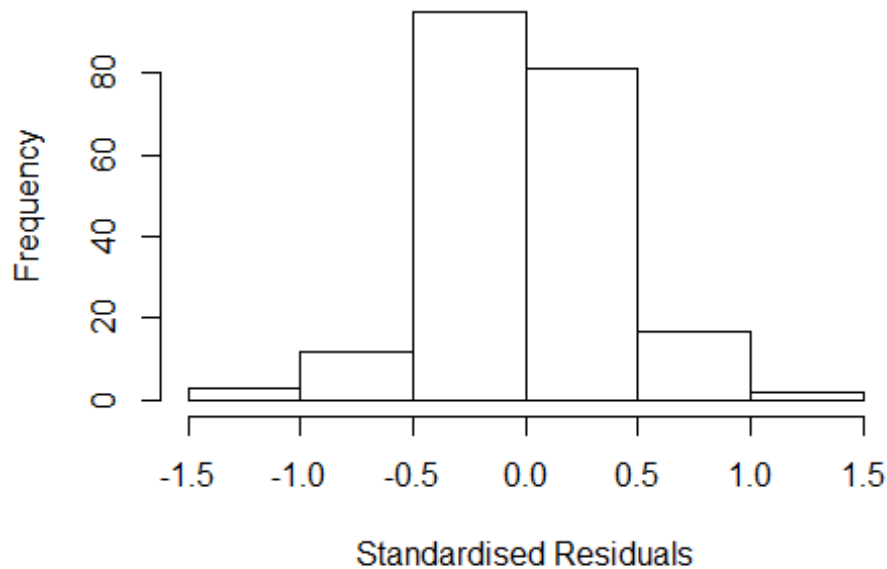
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 73, 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	1.539	1	1.240
LastAuthorFemale	1.802	1	1.342
UniqueAuthors	4.184	4	1.196
Year	7.865	16	1.067

Residuals from first and last author and team size



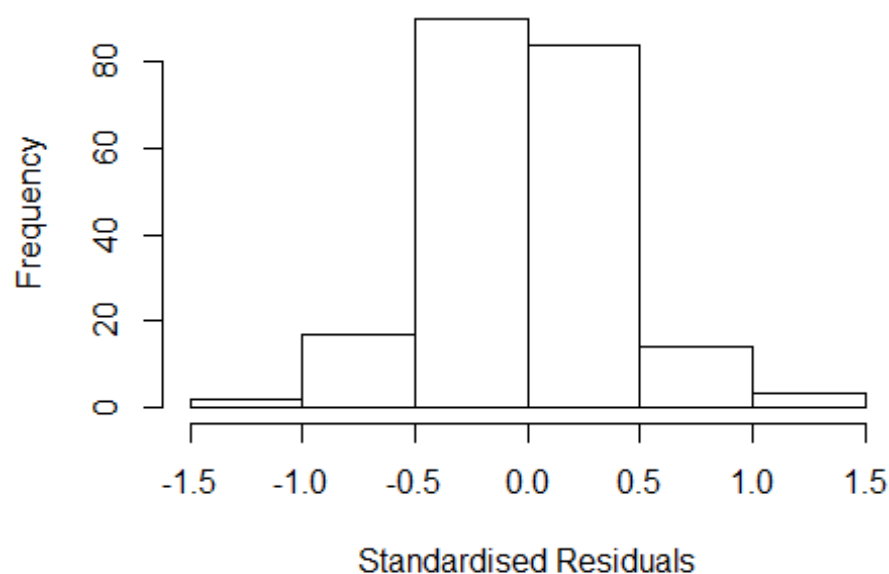
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3664 -0.2309 -0.0115 0.2436 1.3520
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1754 0.1209 9.72 < 2e-16 ***
## FirstAuthorFemale1 -0.0760 0.0765 -0.99 0.322
## LastAuthorFemale1 0.0703 0.0775 0.91 0.365
## UniqueAuthors2 0.1069 0.0872 1.23 0.222
## UniqueAuthors3 0.1770 0.0824 2.15 0.033 *
## UniqueAuthors4 0.0344 0.0970 0.35 0.724
## UniqueAuthors5 0.2059 0.1013 2.03 0.044 *
## Year1997 -0.3177 0.1734 -1.83 0.068 .
## Year1998 0.1972 0.1384 1.43 0.156
## Year1999 -0.1223 0.1466 -0.83 0.405
```

```

## Year2000          1.0132      0.1315      7.71 7.4e-13 ***
## Year2001         -0.0166      0.1540     -0.11  0.914
## Year2002          0.2136      0.1420      1.50  0.134
## Year2003          0.0443      0.1911      0.23  0.817
## Year2004          0.0180      0.1398      0.13  0.898
## Year2005         -0.0196      0.1639     -0.12  0.905
## Year2006         -0.1321      0.1612     -0.82  0.413
## Year2007          0.2027      0.1266      1.60  0.111
## Year2008          0.0841      0.1757      0.48  0.633
## Year2009          0.0565      0.1703      0.33  0.741
## Year2010          0.0361      0.1648      0.22  0.827
## Year2011          0.1041      0.1518      0.69  0.494
## Year2012          0.1128      0.1538      0.73  0.465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.28, Adjusted R-squared:  0.196
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 185 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.875  0.950  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      4.76e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.467 1      1.211
## LastAuthorFemale  1.727 1      1.314
## Year              2.498 16      1.029

```

Residuals from first and last author



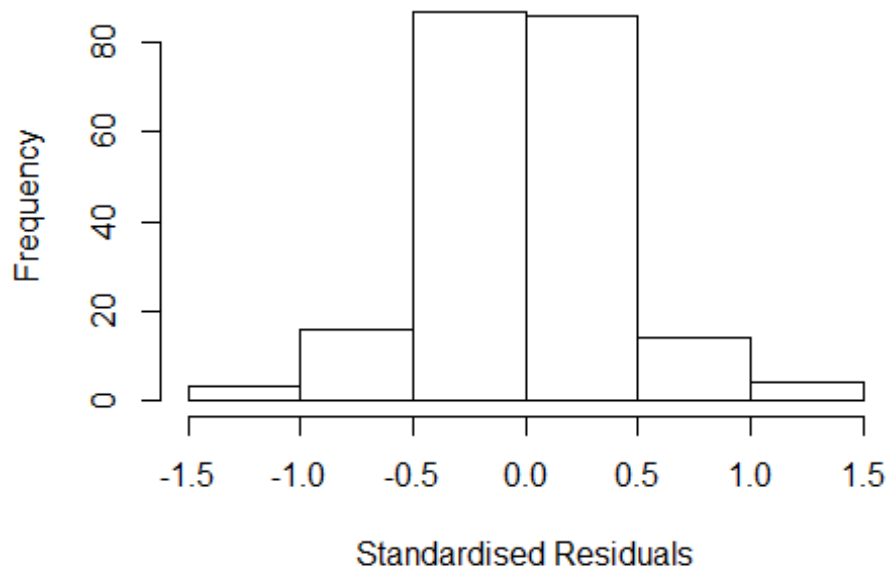
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3621 -0.2145 -0.0253 0.2447 1.2333
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2588 0.1283 9.81 < 2e-16 ***
## FirstAuthorFemale1 -0.0801 0.0747 -1.07 0.285
## LastAuthorFemale1 0.1070 0.0786 1.36 0.175
## Year1997 -0.3557 0.1957 -1.82 0.071 .
## Year1998 0.2002 0.1488 1.35 0.180
## Year1999 -0.1001 0.1647 -0.61 0.544
## Year2000 1.0410 0.1559 6.68 2.6e-10 ***
## Year2001 0.0298 0.1693 0.18 0.861
## Year2002 0.2489 0.1596 1.56 0.121
## Year2003 0.0584 0.2278 0.26 0.798
## Year2004 0.0239 0.1594 0.15 0.881
## Year2005 -0.0151 0.1703 -0.09 0.930
```

```

## Year2006          -0.0947      0.1652    -0.57      0.567
## Year2007           0.2080      0.1436      1.45      0.149
## Year2008           0.0590      0.1879      0.31      0.754
## Year2009           0.1033      0.1802      0.57      0.567
## Year2010           0.0728      0.1682      0.43      0.666
## Year2011           0.1121      0.1837      0.61      0.543
## Year2012           0.1158      0.1584      0.73      0.465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.244, Adjusted R-squared:  0.173
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 195 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.161  0.882  0.956  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      4.76e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.436 1      1.199
## Year              1.436 16      1.011

```


Residuals from first author



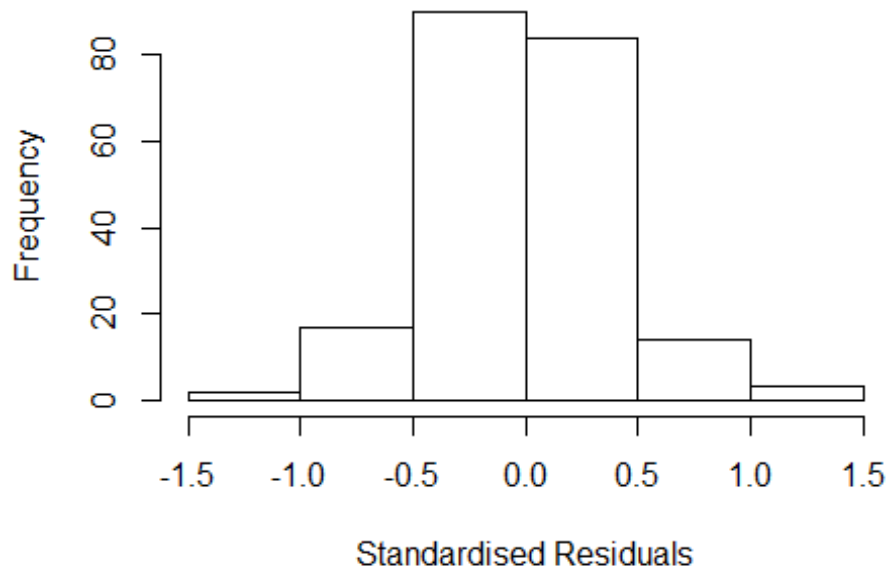
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37291 -0.22659 -0.00487  0.24382  1.22672
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2714     0.1257   10.11 < 2e-16 ***
## FirstAuthorFemale1 -0.0698     0.0744   -0.94  0.350
## Year1997         -0.3709     0.1942   -1.91  0.058 .
## Year1998          0.1896     0.1465    1.29  0.197
## Year1999         -0.0990     0.1692   -0.59  0.559
## Year2000          1.0265     0.1536    6.68 2.5e-10 ***
## Year2001          0.0276     0.1645    0.17  0.867
## Year2002          0.2428     0.1569    1.55  0.123
## Year2003          0.0844     0.2390    0.35  0.724
## Year2004          0.0188     0.1578    0.12  0.905
## Year2005          0.0094     0.1694    0.06  0.956
## Year2006         -0.1096     0.1629   -0.67  0.502
```

```

## Year2007          0.2132      0.1440      1.48      0.140
## Year2008          0.0869      0.1842      0.47      0.638
## Year2009          0.1015      0.1822      0.56      0.578
## Year2010          0.0718      0.1681      0.43      0.669
## Year2011          0.0994      0.1818      0.55      0.585
## Year2012          0.1133      0.1573      0.72      0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.237, Adjusted R-squared:  0.169
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 190 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.157  0.875   0.954   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      4.76e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.813 1      1.346
## Year              1.813 16      1.019

```

Residuals from last author



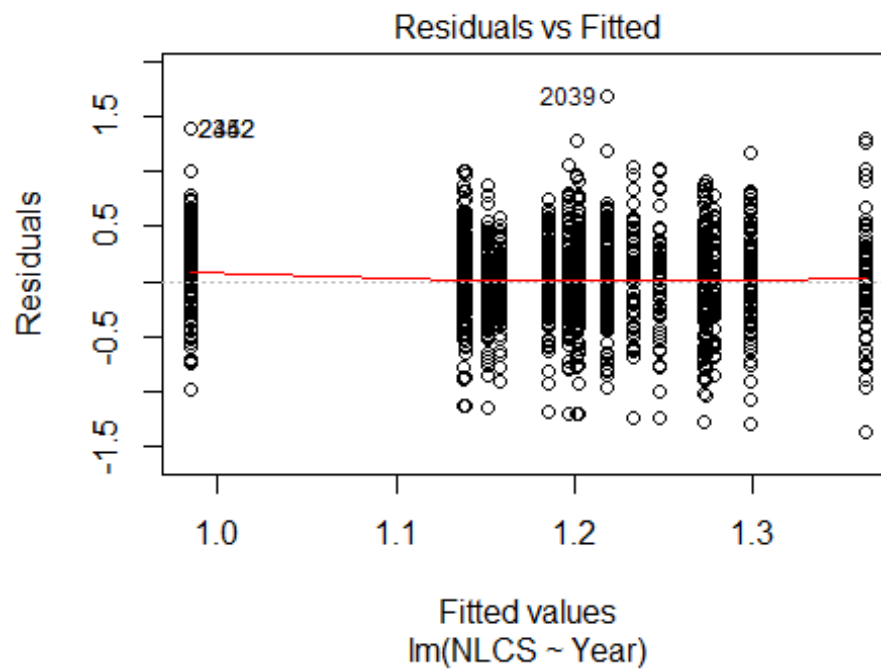
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3505 -0.2250 -0.0245 0.2445 1.2331
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2598 0.1284 9.81 < 2e-16 ***
## LastAuthorFemale1 0.0992 0.0821 1.21 0.228
## Year1997 -0.3754 0.1978 -1.90 0.059 .
## Year1998 0.1873 0.1478 1.27 0.207
## Year1999 -0.1130 0.1654 -0.68 0.495
## Year2000 1.0255 0.1553 6.60 3.8e-10 ***
## Year2001 0.0213 0.1716 0.12 0.901
## Year2002 0.2481 0.1596 1.55 0.122
## Year2003 0.0402 0.2549 0.16 0.875
## Year2004 0.0147 0.1587 0.09 0.926
## Year2005 -0.0345 0.1698 -0.20 0.839
## Year2006 -0.1140 0.1642 -0.69 0.488
```

```

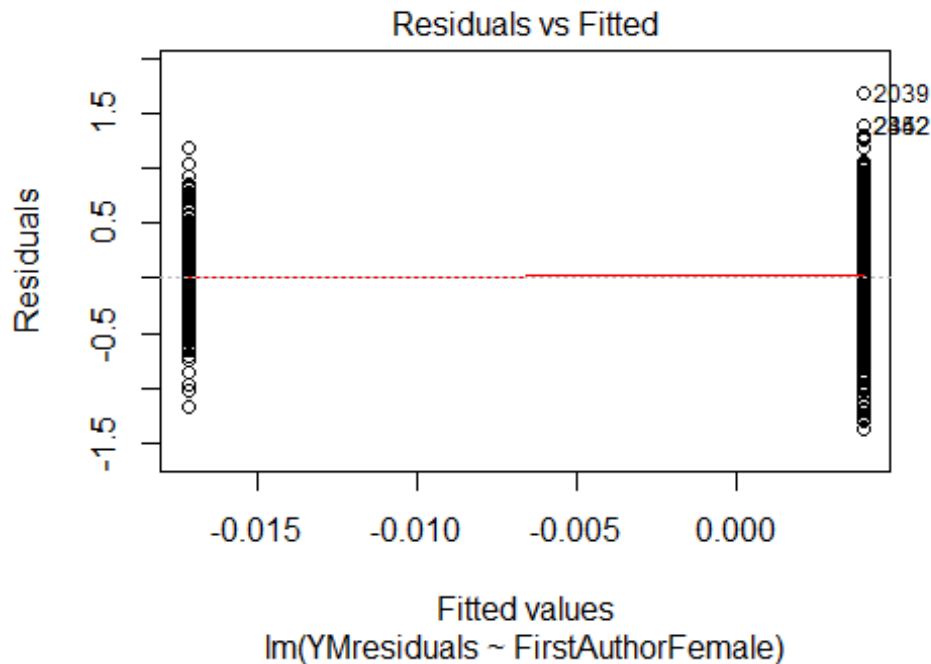
## Year2007          0.2035      0.1434      1.42      0.157
## Year2008          0.0450      0.1850      0.24      0.808
## Year2009          0.0907      0.1845      0.49      0.624
## Year2010          0.0726      0.1683      0.43      0.667
## Year2011          0.1111      0.1838      0.60      0.546
## Year2012          0.0959      0.1572      0.61      0.542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.24,   Adjusted R-squared:  0.173
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 199 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.166  0.878  0.959  0.905  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      4.76e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 210"
## [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
##  161  150  215  151  199  206  215  174  227  204  206  206  212  200  205
## 2011 2012
##  190  205
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   65   59   99   61   58   69  103   86  128  131  120  134  113  119  105
## 2011 2012

```

```
## 97 112
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62 54 90 55 49 62 93 78 110 105 98 108 99 104 96
## 2011 2012
## 81 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 = 81, df = 16, p-value = 1e-10
```

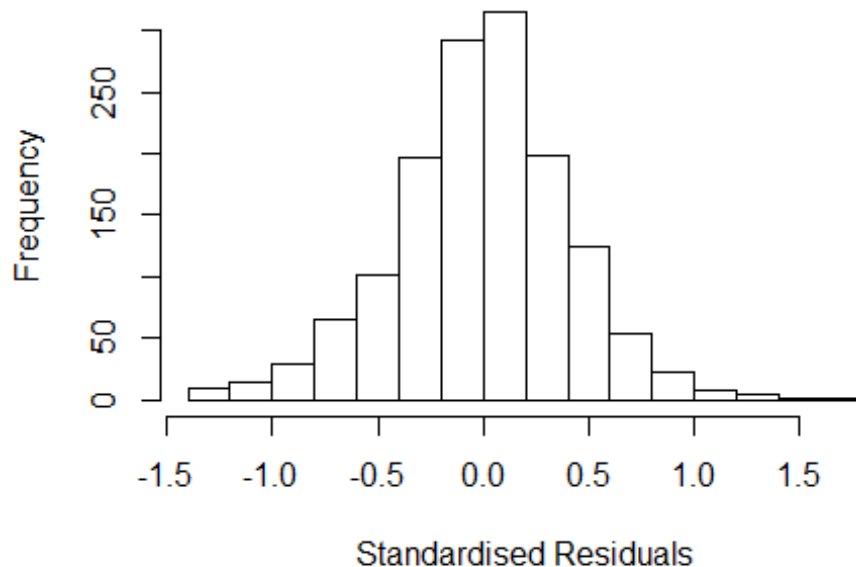


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.85286788584336"
## [1] "Male first author team size 2018 geometric mean: 3.03536333299365"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.53003577953455"
## [1] "Male last author team size 2018 geometric mean: 3.21224707938736"
##
## 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] "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  1.081 1      1.040
## UniqueAuthors    1.347 4      1.038
## Year              1.460 16     1.012
```

Residuals from first and last author and team size



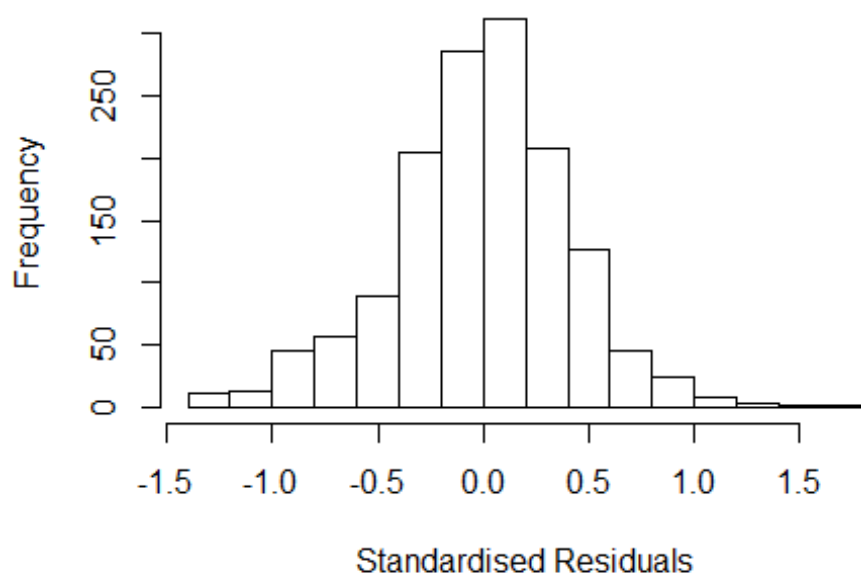
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29063 -0.24621  0.00622  0.24845  1.66719
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22485    0.06908   17.73 < 2e-16 ***
## FirstAuthorFemale1 -0.06394    0.02706   -2.36  0.01828 *
## LastAuthorFemale1 -0.01234    0.02978   -0.41  0.67853
## UniqueAuthors2     0.10791    0.03645    2.96  0.00313 **
## UniqueAuthors3     0.16253    0.03766    4.32  1.7e-05 ***
## UniqueAuthors4     0.14222    0.03973    3.58  0.00036 ***
## UniqueAuthors5     0.22901    0.04310    5.31  1.2e-07 ***
## Year1997          -0.06333    0.08515   -0.74  0.45716
## Year1998          -0.02137    0.07654   -0.28  0.78009
## Year1999          -0.05886    0.08446   -0.70  0.48596
```

```

## Year2000          0.04108      0.11122      0.37  0.71193
## Year2001          -0.07060      0.08238     -0.86  0.39155
## Year2002           0.00523      0.07796      0.07  0.94653
## Year2003          -0.17420      0.08001     -2.18  0.02963 *
## Year2004          -0.11891      0.07308     -1.63  0.10395
## Year2005          -0.39713      0.08404     -4.73  2.5e-06 ***
## Year2006          -0.16435      0.07212     -2.28  0.02283 *
## Year2007          -0.10675      0.07267     -1.47  0.14204
## Year2008          -0.14344      0.06767     -2.12  0.03421 *
## Year2009          -0.18352      0.07153     -2.57  0.01040 *
## Year2010          -0.16323      0.07447     -2.19  0.02855 *
## Year2011          -0.19014      0.07452     -2.55  0.01083 *
## Year2012          -0.14363      0.07455     -1.93  0.05423 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.0852, Adjusted R-squared:  0.071
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1309 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0088 0.8560 0.9520 0.8880 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          6.95e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1          1.032
## LastAuthorFemale  1.085 1          1.042
## Year              1.148 16          1.004

```


Residuals from first and last author



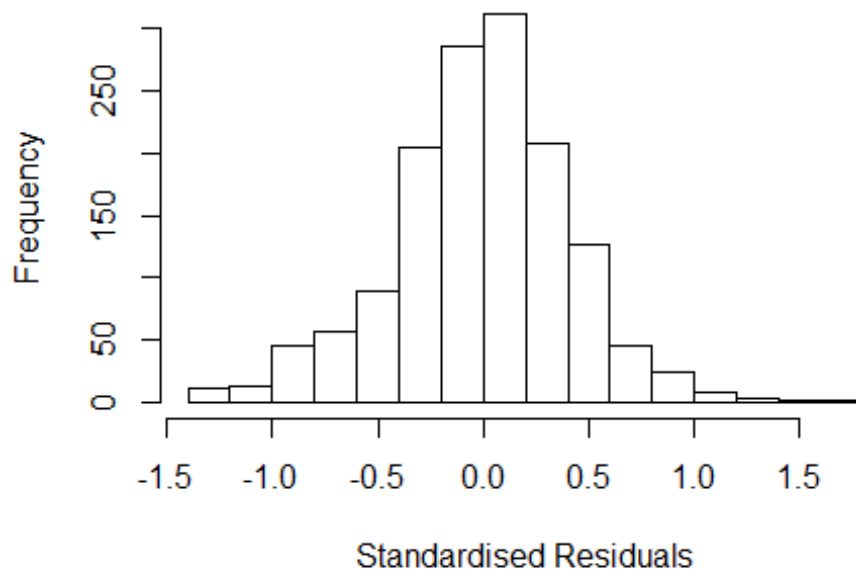
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36447 -0.24775  0.00614  0.25339  1.69196
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.320153   0.062872   21.00  < 2e-16 ***
## FirstAuthorFemale1 -0.050692   0.027494   -1.84   0.065 .
## LastAuthorFemale1 -0.000508   0.030764   -0.02   0.987
## Year1997        -0.061890   0.086391   -0.72   0.474
## Year1998        -0.028404   0.077735   -0.37   0.715
## Year1999        -0.054431   0.084119   -0.65   0.518
## Year2000         0.044319   0.111617    0.40   0.691
## Year2001        -0.053406   0.081619   -0.65   0.513
## Year2002         0.031520   0.078111    0.40   0.687
## Year2003        -0.149424   0.078358   -1.91   0.057 .
## Year2004        -0.108609   0.073190   -1.48   0.138
## Year2005        -0.372530   0.086572   -4.30  1.8e-05 ***
```

```

## Year2006      -0.145744    0.071593   -2.04    0.042 *
## Year2007      -0.088773    0.071805   -1.24    0.217
## Year2008      -0.116671    0.068252   -1.71    0.088 .
## Year2009      -0.143816    0.071884   -2.00    0.046 *
## Year2010      -0.140600    0.075037   -1.87    0.061 .
## Year2011      -0.152081    0.074587   -2.04    0.042 *
## Year2012      -0.110219    0.074247   -1.48    0.138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.055, Adjusted R-squared:  0.043
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1317 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8560 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      6.95e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



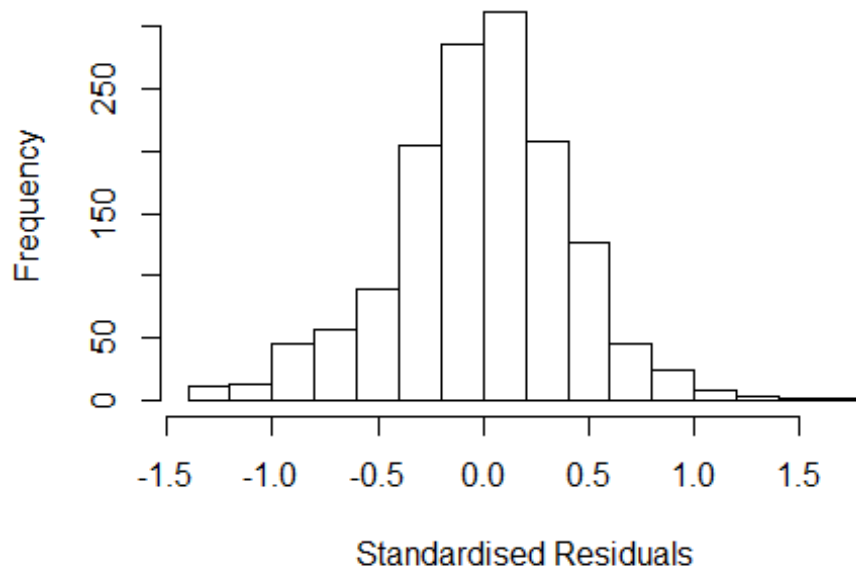
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36442 -0.24770  0.00626  0.25344  1.69152
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3202     0.0627   21.06 < 2e-16 ***
## FirstAuthorFemale1 -0.0508     0.0276   -1.84   0.066 .
## Year1997         -0.0619     0.0864   -0.72   0.474
## Year1998         -0.0285     0.0777   -0.37   0.714
## Year1999         -0.0545     0.0840   -0.65   0.517
## Year2000          0.0443     0.1116    0.40   0.692
## Year2001         -0.0534     0.0816   -0.65   0.513
## Year2002          0.0315     0.0781    0.40   0.687
## Year2003         -0.1495     0.0784   -1.91   0.057 .
## Year2004         -0.1087     0.0732   -1.48   0.138
## Year2005         -0.3725     0.0864   -4.31 1.7e-05 ***
## Year2006         -0.1458     0.0716   -2.04   0.042 *
```

```

## Year2007          -0.0888      0.0718   -1.24    0.216
## Year2008          -0.1167      0.0683   -1.71    0.088 .
## Year2009          -0.1439      0.0719   -2.00    0.046 *
## Year2010          -0.1406      0.0751   -1.87    0.061 .
## Year2011          -0.1522      0.0747   -2.04    0.042 *
## Year2012          -0.1103      0.0743   -1.48    0.138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0549, Adjusted R-squared:  0.0436
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1317 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8560 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      6.95e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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

```

Residuals from last author



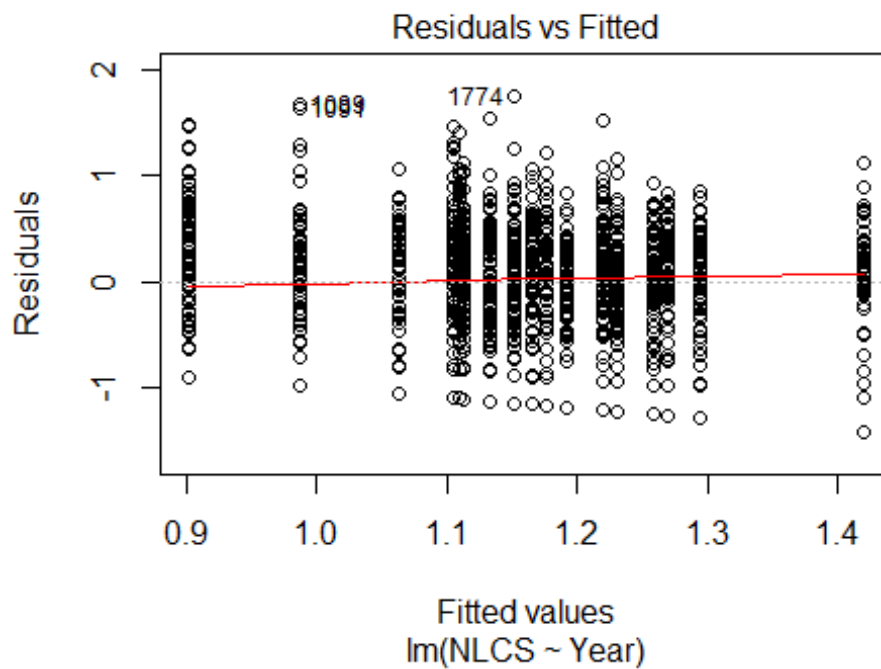
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36188 -0.25195 0.00629 0.25000 1.70770
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31662 0.06276 20.98 < 2e-16 ***
## LastAuthorFemale1 -0.00904 0.03134 -0.29 0.773
## Year1997 -0.06241 0.08654 -0.72 0.471
## Year1998 -0.02947 0.07763 -0.38 0.704
## Year1999 -0.06107 0.08353 -0.73 0.465
## Year2000 0.04526 0.11255 0.40 0.688
## Year2001 -0.05430 0.08197 -0.66 0.508
## Year2002 0.02763 0.07831 0.35 0.724
## Year2003 -0.15856 0.07878 -2.01 0.044 *
## Year2004 -0.11229 0.07325 -1.53 0.126
## Year2005 -0.37535 0.08686 -4.32 1.7e-05 ***
## Year2006 -0.15394 0.07154 -2.15 0.032 *
```

```

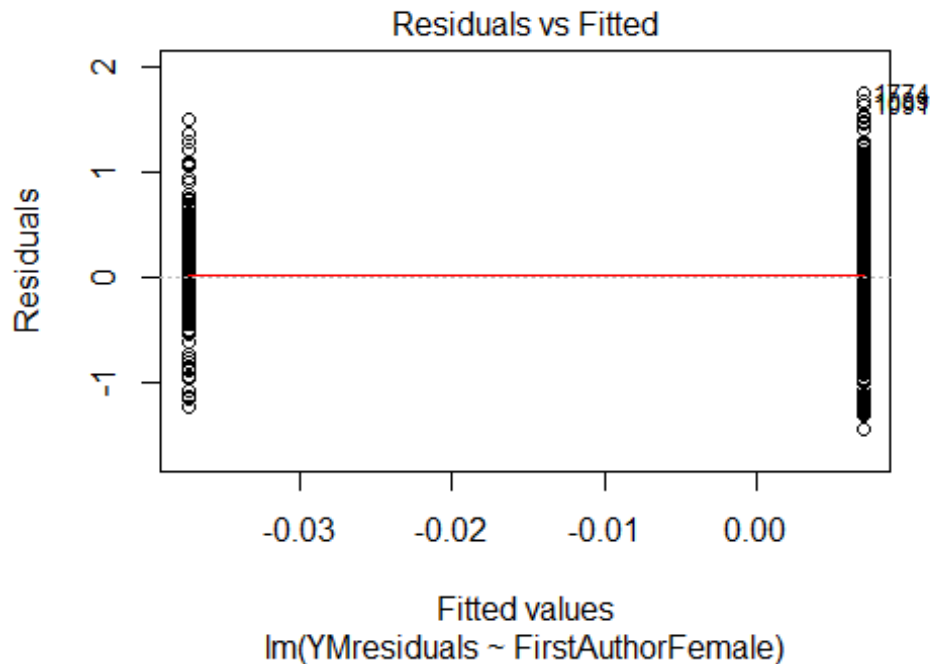
## Year2007          -0.09198      0.07161    -1.28      0.199
## Year2008          -0.12018      0.06826    -1.76      0.079 .
## Year2009          -0.15492      0.07150    -2.17      0.030 *
## Year2010          -0.14639      0.07510    -1.95      0.051 .
## Year2011          -0.15932      0.07450    -2.14      0.033 *
## Year2012          -0.11596      0.07416    -1.56      0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.0522, Adjusted R-squared:  0.0408
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1318 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0043 0.8580 0.9500 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      6.95e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1439"
## [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
## 129 175 212 140 220 174 177 109 180 170 143 153 192 148 169
## 2011 2012
## 173 150
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 65 111 64 89 73 100 67 112 120 73 96 107 88 93
## 2011 2012

```

```
##      88      87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   49   57   97   59   82   62   91   61  107  101   61   82   94   84   87
## 2011 2012
##   78   78
## [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
```

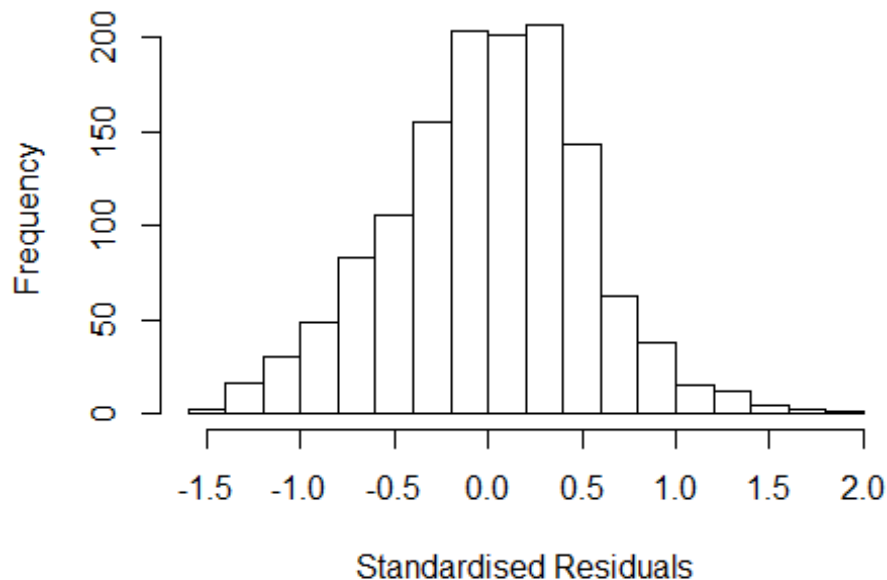


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.81633977249117"
## [1] "Male first author team size 2018 geometric mean: 3.65697839344512"
##
## 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] "Female last author team size 2018 geometric mean: 3.56095110548341"
## [1] "Male last author team size 2018 geometric mean: 3.76101107640751"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.148 1      1.072
## LastAuthorFemale  1.157 1      1.075
## UniqueAuthors    1.312 4      1.035
## Year              1.480 16     1.012
```


Residuals from first and last author and team size



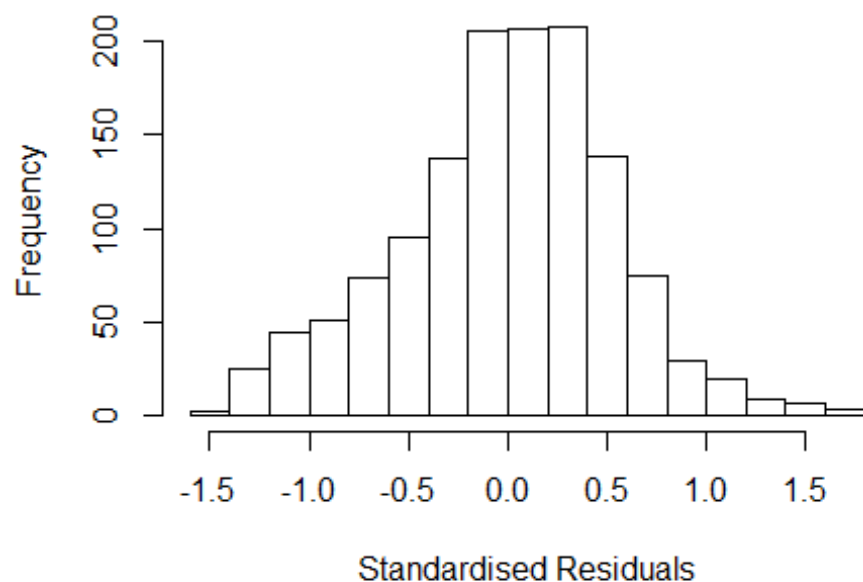
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5260 -0.3242 0.0171 0.3417 1.8265
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2812 0.0792 16.18 < 2e-16 ***
## FirstAuthorFemale1 -0.0550 0.0378 -1.46 0.14562
## LastAuthorFemale1 -0.0211 0.0418 -0.50 0.61403
## UniqueAuthors2 0.1948 0.0448 4.35 1.5e-05 ***
## UniqueAuthors3 0.3115 0.0447 6.97 5.0e-12 ***
## UniqueAuthors4 0.3356 0.0514 6.53 9.7e-11 ***
## UniqueAuthors5 0.3722 0.0486 7.65 3.8e-14 ***
## Year1997 -0.2881 0.1030 -2.80 0.00524 **
## Year1998 -0.3019 0.1037 -2.91 0.00367 **
## Year1999 -0.1274 0.0978 -1.30 0.19283
```

```

## Year2000      -0.5608      0.1022      -5.49      4.9e-08 ***
## Year2001      -0.4049      0.1177      -3.44      0.00060 ***
## Year2002      -0.2178      0.0994      -2.19      0.02861 *
## Year2003      -0.3790      0.1004      -3.78      0.00017 ***
## Year2004      -0.2867      0.0868      -3.30      0.00098 ***
## Year2005      -0.6068      0.1000      -6.07      1.7e-09 ***
## Year2006      -0.2778      0.1052      -2.64      0.00837 **
## Year2007      -0.2175      0.0939      -2.32      0.02065 *
## Year2008      -0.3002      0.0853      -3.52      0.00045 ***
## Year2009      -0.3212      0.0866      -3.71      0.00022 ***
## Year2010      -0.3568      0.0905      -3.94      8.4e-05 ***
## Year2011      -0.4227      0.0964      -4.38      1.3e-05 ***
## Year2012      -0.2130      0.0901      -2.36      0.01822 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.49
## Multiple R-squared:  0.128, Adjusted R-squared:  0.114
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 101 weights are ~= 1. The remaining 1229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.135  0.866  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          7.52e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.117 1 1.057
## LastAuthorFemale 1.119 1 1.058
## Year 1.140 16 1.004

```

Residuals from first and last author



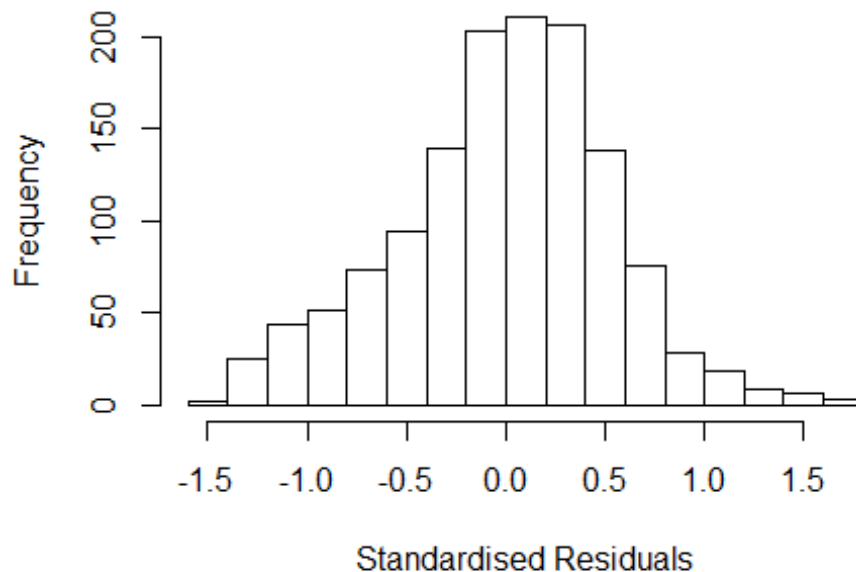
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4338 -0.3303  0.0252  0.3377  1.7436
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.43375    0.07763   18.47  < 2e-16 ***
## FirstAuthorFemale1 -0.03358    0.03955   -0.85  0.39609
## LastAuthorFemale1 -0.00991    0.04339   -0.23  0.81947
## Year1997        -0.24820    0.10580   -2.35  0.01912 *
## Year1998        -0.29374    0.10732   -2.74  0.00628 **
## Year1999        -0.11091    0.10331   -1.07  0.28319
## Year2000        -0.50482    0.10054   -5.02  5.8e-07 ***
## Year2001        -0.38024    0.12350   -3.08  0.00212 **
## Year2002        -0.13698    0.10195   -1.34  0.17930
## Year2003        -0.33268    0.10778   -3.09  0.00207 **
## Year2004        -0.26444    0.09166   -2.88  0.00398 **
## Year2005        -0.59992    0.10733   -5.59  2.8e-08 ***
```

```

## Year2006      -0.22648    0.10813   -2.09  0.03641 *
## Year2007      -0.15360    0.09512   -1.61  0.10661
## Year2008      -0.20024    0.08746   -2.29  0.02221 *
## Year2009      -0.21563    0.08807   -2.45  0.01448 *
## Year2010      -0.28913    0.09268   -3.12  0.00185 **
## Year2011      -0.34459    0.10045   -3.43  0.00062 ***
## Year2012      -0.11467    0.09136   -1.26  0.20965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0694, Adjusted R-squared:  0.0566
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 103 weights are ~= 1. The remaining 1227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.853   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      7.52e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"    "nonsingular"      ".vcov.avar1"
## 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



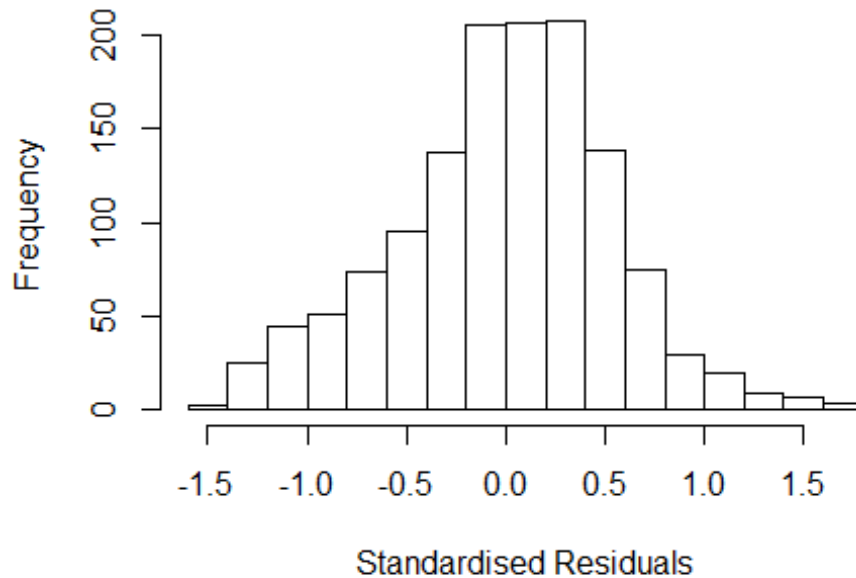
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4336 -0.3296 0.0265 0.3356 1.7344
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4336 0.0777 18.45 < 2e-16 ***
## FirstAuthorFemale1 -0.0358 0.0388 -0.92 0.35588
## Year1997 -0.2482 0.1058 -2.35 0.01913 *
## Year1998 -0.2942 0.1073 -2.74 0.00621 **
## Year1999 -0.1110 0.1033 -1.07 0.28271
## Year2000 -0.5054 0.1005 -5.03 5.7e-07 ***
## Year2001 -0.3808 0.1234 -3.09 0.00207 **
## Year2002 -0.1370 0.1020 -1.34 0.17965
## Year2003 -0.3340 0.1074 -3.11 0.00190 **
## Year2004 -0.2650 0.0916 -2.89 0.00390 **
## Year2005 -0.6015 0.1076 -5.59 2.7e-08 ***
## Year2006 -0.2265 0.1082 -2.09 0.03652 *
```

```

## Year2007          -0.1542      0.0950   -1.62   0.10475
## Year2008          -0.2020      0.0871   -2.32   0.02053 *
## Year2009          -0.2172      0.0877   -2.48   0.01342 *
## Year2010          -0.2901      0.0927   -3.13   0.00179 **
## Year2011          -0.3457      0.1003   -3.45   0.00058 ***
## Year2012          -0.1160      0.0911   -1.27   0.20272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0694, Adjusted R-squared:  0.0573
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1230 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.196  0.853  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      7.52e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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
## -1.4324 -0.3280 0.0273 0.3367 1.7559
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4324 0.0774 18.51 < 2e-16 ***
## LastAuthorFemale1 -0.0184 0.0428 -0.43 0.66686
## Year1997 -0.2502 0.1061 -2.36 0.01854 *
## Year1998 -0.2955 0.1073 -2.75 0.00598 **
## Year1999 -0.1130 0.1033 -1.09 0.27396
## Year2000 -0.5043 0.1004 -5.03 5.7e-07 ***
## Year2001 -0.3805 0.1234 -3.08 0.00209 **
## Year2002 -0.1410 0.1012 -1.39 0.16381
## Year2003 -0.3336 0.1081 -3.09 0.00206 **
## Year2004 -0.2669 0.0914 -2.92 0.00356 **
## Year2005 -0.6029 0.1076 -5.60 2.6e-08 ***
## Year2006 -0.2290 0.1079 -2.12 0.03390 *
```

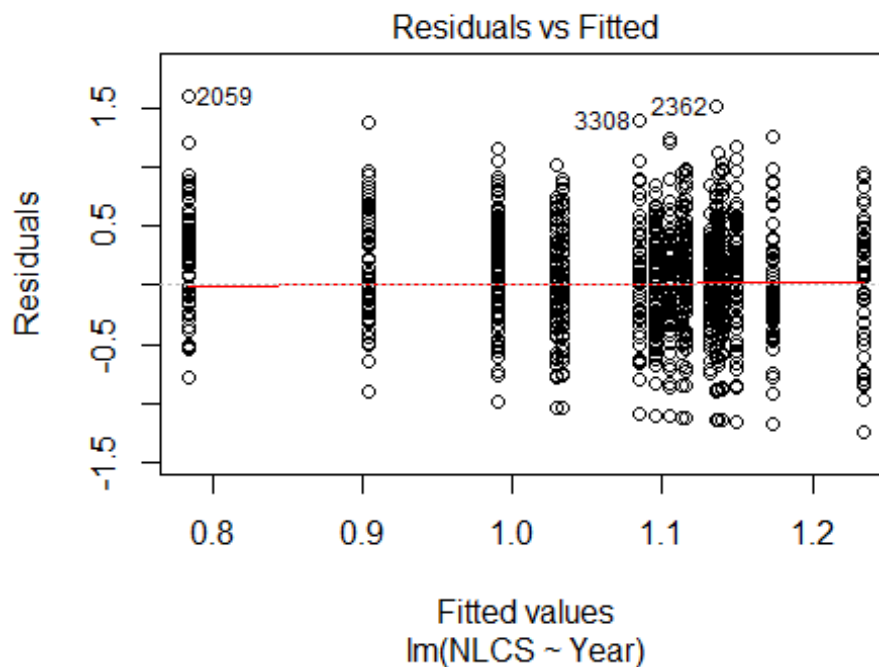
```

## Year2007          -0.1581      0.0950   -1.66   0.09640 .
## Year2008          -0.2025      0.0872   -2.32   0.02042 *
## Year2009          -0.2196      0.0879   -2.50   0.01264 *
## Year2010          -0.2926      0.0926   -3.16   0.00160 **
## Year2011          -0.3490      0.0999   -3.49   0.00049 ***
## Year2012          -0.1200      0.0909   -1.32   0.18721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0688, Adjusted R-squared:  0.0568
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1223 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.184  0.854  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      7.52e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1330"
## [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
## 122 150 224 121 198 158 154 141 172 145 149 128 123 112 121
## 2011 2012
## 120 155
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 61 117 57 90 68 83 79 98 106 91 94 71 61 74
## 2011 2012

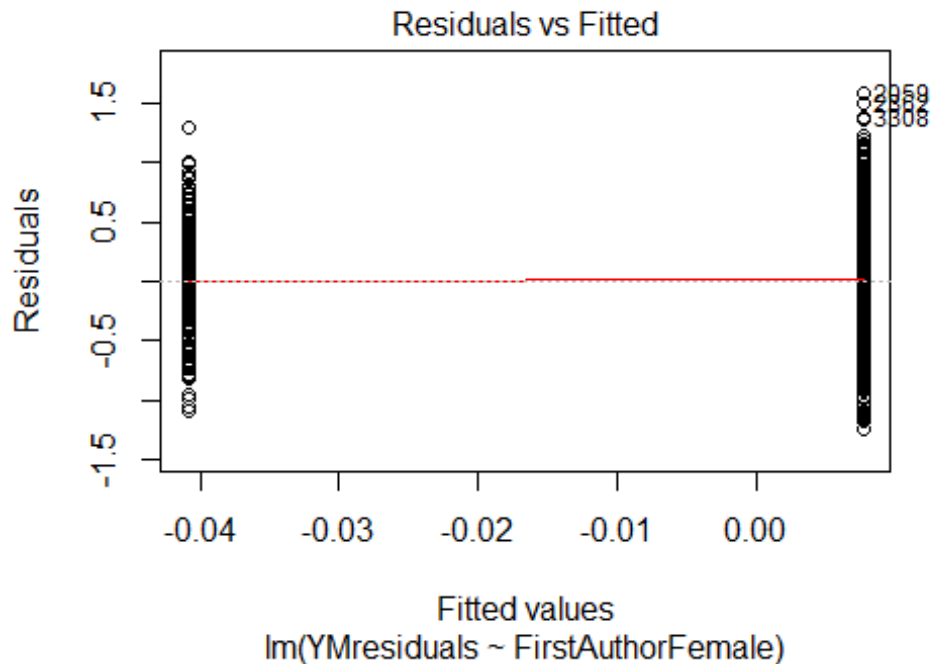
```



```
## 62 82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 55 105 53 83 63 76 73 88 88 75 73 56 56 69
## 2011 2012
## 56 77
## [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
```



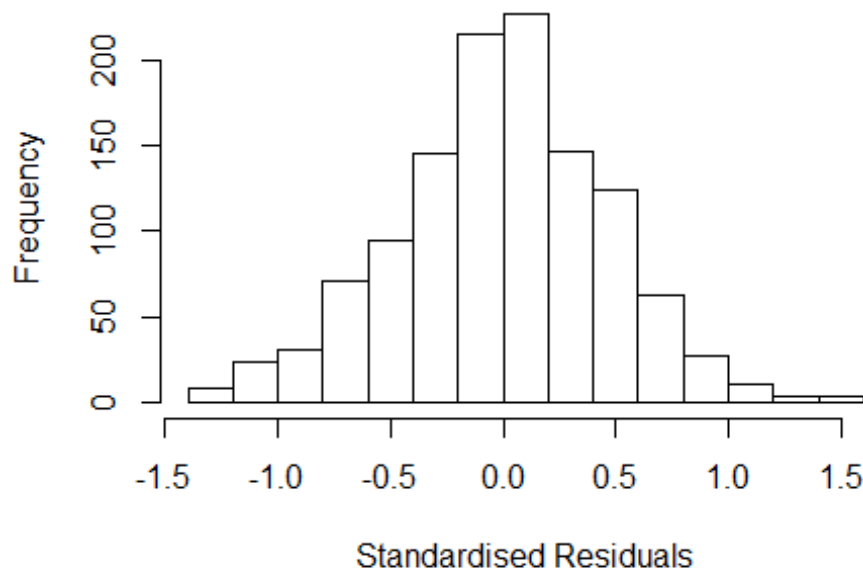
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.73, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 3.35433616388843"
## [1] "Male first author team size 2018 geometric mean: 2.87792587965069"
## 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 = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.66716827534"
## [1] "Male last author team size 2018 geometric mean: 3.06897124910951"
##
## 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	1.120	1	1.059
LastAuthorFemale	1.098	1	1.048
UniqueAuthors	1.344	4	1.038
Year	1.535	16	1.013

Residuals from first and last author and team size



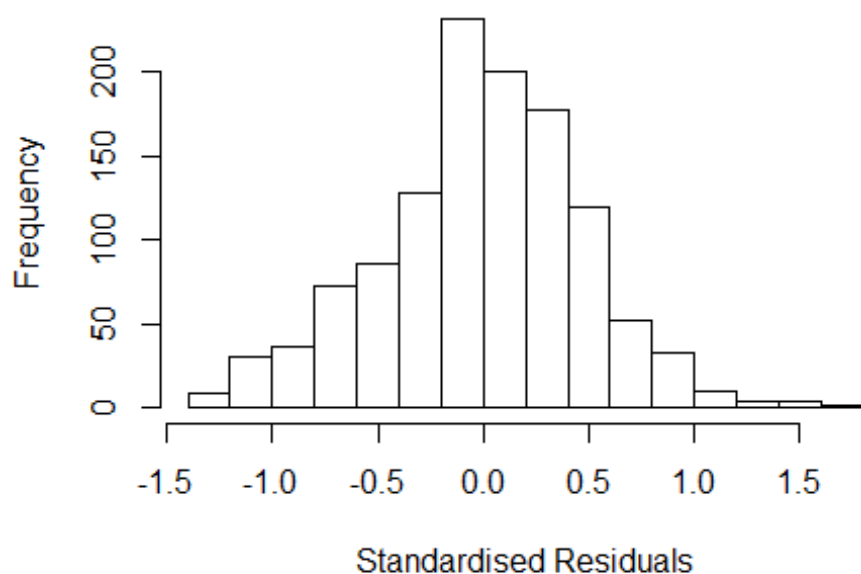
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3553 -0.2891 0.0107 0.2994 1.5609
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1798 0.0870 13.56 < 2e-16 ***
## FirstAuthorFemale1 -0.0966 0.0358 -2.70 0.0071 **
## LastAuthorFemale1 0.0483 0.0421 1.15 0.2511
## UniqueAuthors2 0.0989 0.0398 2.48 0.0132 *
## UniqueAuthors3 0.2392 0.0416 5.75 1.1e-08 ***
## UniqueAuthors4 0.2787 0.0473 5.90 4.9e-09 ***
## UniqueAuthors5 0.3036 0.0532 5.71 1.5e-08 ***
## Year1997 -0.2420 0.1055 -2.29 0.0220 *
## Year1998 -0.2691 0.0975 -2.76 0.0059 **
## Year1999 -0.1281 0.1075 -1.19 0.2336
```

```

## Year2000          -0.4547      0.1014    -4.48    8.1e-06 ***
## Year2001          -0.1552      0.1039    -1.49     0.1356
## Year2002          -0.1311      0.1024    -1.28     0.2005
## Year2003          -0.2673      0.0990    -2.70     0.0070 **
## Year2004          -0.2099      0.0914    -2.30     0.0217 *
## Year2005          -0.5742      0.1026    -5.60    2.7e-08 ***
## Year2006          -0.1876      0.0953    -1.97     0.0494 *
## Year2007          -0.2125      0.0961    -2.21     0.0272 *
## Year2008          -0.1710      0.0914    -1.87     0.0614 .
## Year2009          -0.1717      0.1030    -1.67     0.0957 .
## Year2010          -0.1502      0.1001    -1.50     0.1339
## Year2011          -0.3048      0.0997    -3.06     0.0023 **
## Year2012          -0.1905      0.0984    -1.93     0.0532 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.129, Adjusted R-squared:  0.113
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1084 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.156  0.852  0.947   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      8.39e-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 1.097 1      1.048
## LastAuthorFemale  1.104 1      1.051
## Year              1.192 16      1.006

```

Residuals from first and last author



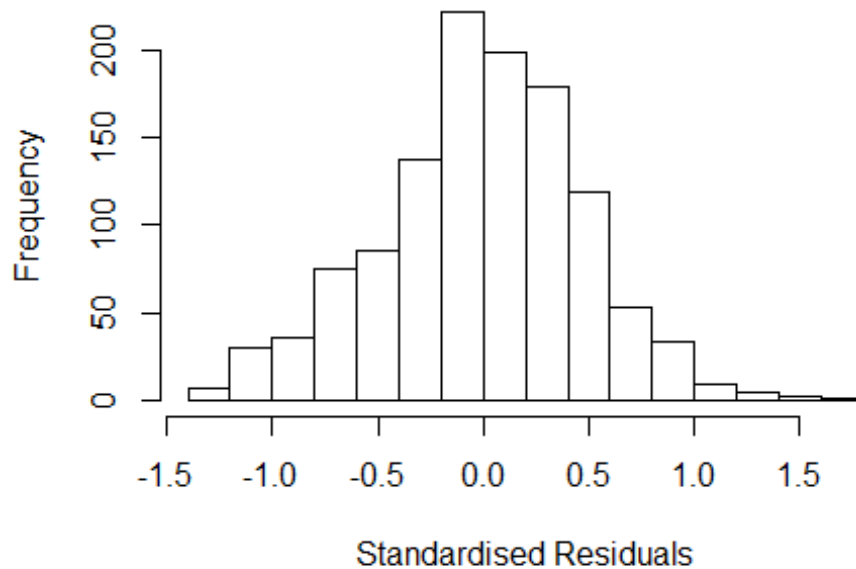
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29458 -0.28226  0.00332  0.30042  1.64430
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2946    0.0870   14.87 < 2e-16 ***
## FirstAuthorFemale1 -0.0826    0.0373   -2.22  0.0268 *
## LastAuthorFemale1  0.0551    0.0450    1.22  0.2215
## Year1997          -0.2466    0.1088   -2.27  0.0236 *
## Year1998          -0.2761    0.1026   -2.69  0.0072 **
## Year1999          -0.1278    0.1122   -1.14  0.2549
## Year2000          -0.4223    0.1049   -4.03 6.0e-05 ***
## Year2001          -0.1387    0.1063   -1.31  0.1921
## Year2002          -0.0908    0.1063   -0.85  0.3930
## Year2003          -0.2388    0.1026   -2.33  0.0201 *
## Year2004          -0.1997    0.0975   -2.05  0.0408 *
## Year2005          -0.5609    0.1101   -5.09 4.1e-07 ***
```

```

## Year2006          -0.1523      0.0977    -1.56    0.1191
## Year2007          -0.1831      0.0996    -1.84    0.0663 .
## Year2008          -0.1279      0.0953    -1.34    0.1797
## Year2009          -0.1222      0.1057    -1.16    0.2480
## Year2010          -0.1213      0.1024    -1.18    0.2366
## Year2011          -0.2320      0.1073    -2.16    0.0308 *
## Year2012          -0.1298      0.1022    -1.27    0.2044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0771, Adjusted R-squared:  0.063
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1093 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.131  0.850  0.947  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      8.39e-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 1.091 1      1.044
## Year              1.091 16      1.003

```

Residuals from first author



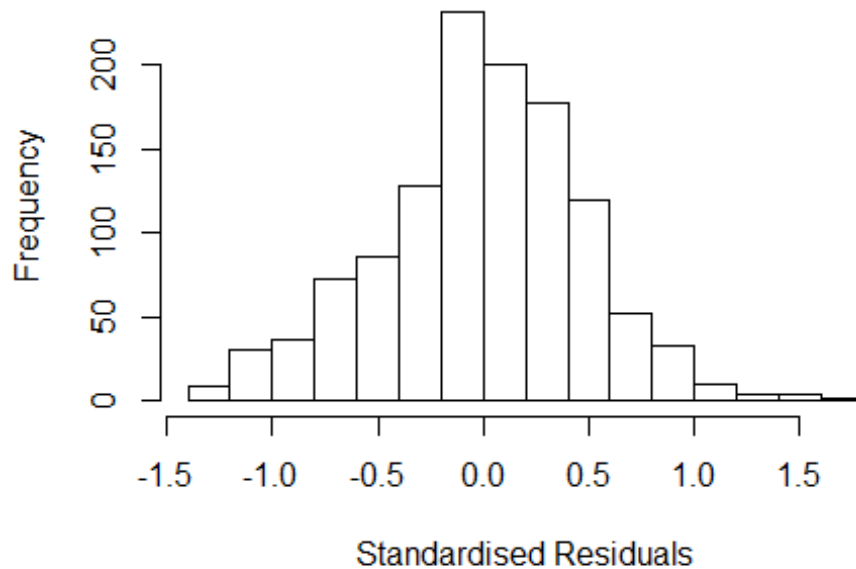
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30219 -0.28160 0.00343 0.30749 1.64043
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3022 0.0874 14.89 < 2e-16 ***
## FirstAuthorFemale1 -0.0728 0.0370 -1.96 0.0497 *
## Year1997 -0.2517 0.1093 -2.30 0.0215 *
## Year1998 -0.2800 0.1034 -2.71 0.0069 **
## Year1999 -0.1356 0.1121 -1.21 0.2269
## Year2000 -0.4251 0.1057 -4.02 6.1e-05 ***
## Year2001 -0.1454 0.1069 -1.36 0.1740
## Year2002 -0.0948 0.1071 -0.89 0.3760
## Year2003 -0.2409 0.1034 -2.33 0.0199 *
## Year2004 -0.2019 0.0985 -2.05 0.0406 *
## Year2005 -0.5646 0.1094 -5.16 2.9e-07 ***
## Year2006 -0.1563 0.0984 -1.59 0.1126
```

```

## Year2007          -0.1873      0.1002    -1.87    0.0619 .
## Year2008          -0.1309      0.0962    -1.36    0.1738
## Year2009          -0.1241      0.1064    -1.17    0.2437
## Year2010          -0.1237      0.1033    -1.20    0.2313
## Year2011          -0.2342      0.1081    -2.17    0.0304 *
## Year2012          -0.1336      0.1028    -1.30    0.1940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.076, Adjusted R-squared:  0.0626
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1104 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.852  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      8.39e-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 1.118 1          1.057
## Year            1.118 16          1.003

```


Residuals from last author



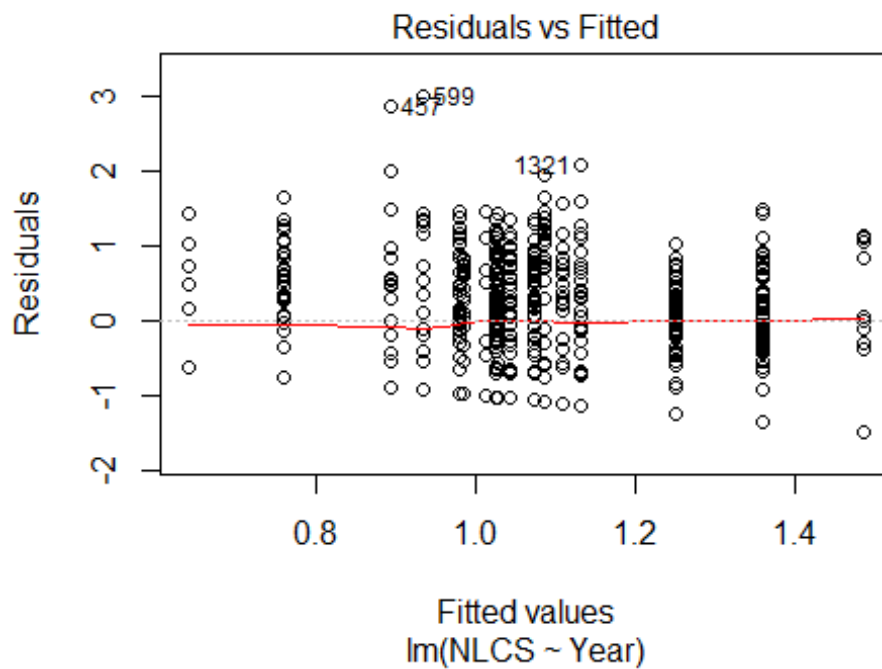
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29242 -0.28464 0.00368 0.30478 1.65427
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2924 0.0873 14.81 < 2e-16 ***
## LastAuthorFemale1 0.0354 0.0471 0.75 0.4522
## Year1997 -0.2525 0.1091 -2.31 0.0208 *
## Year1998 -0.2805 0.1030 -2.72 0.0066 **
## Year1999 -0.1388 0.1116 -1.24 0.2139
## Year2000 -0.4266 0.1051 -4.06 5.2e-05 ***
## Year2001 -0.1435 0.1065 -1.35 0.1779
## Year2002 -0.0986 0.1064 -0.93 0.3541
## Year2003 -0.2524 0.1031 -2.45 0.0145 *
## Year2004 -0.2050 0.0978 -2.10 0.0363 *
## Year2005 -0.5687 0.1114 -5.10 3.9e-07 ***
## Year2006 -0.1726 0.0974 -1.77 0.0767 .
```

```

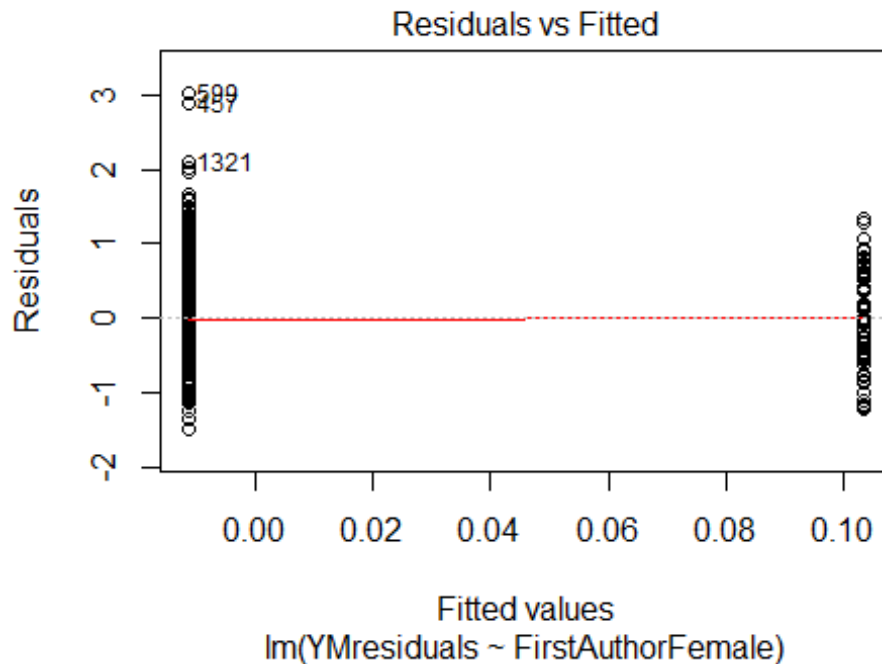
## Year2007          -0.1889      0.1002    -1.89    0.0595 .
## Year2008          -0.1381      0.0957    -1.44    0.1493
## Year2009          -0.1375      0.1058    -1.30    0.1937
## Year2010          -0.1377      0.1027    -1.34    0.1804
## Year2011          -0.2408      0.1070    -2.25    0.0247 *
## Year2012          -0.1394      0.1032    -1.35    0.1770
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0733, Adjusted R-squared:  0.0598
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1093 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.125  0.846   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      8.39e-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: 1192"
## [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
##   59   85   74   98  114  198  129  127  107  121  136  137  129  148  186
## 2011 2012
##  177  202
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   11   12    9   24   27   32   49   46   33   39   54   64   42   52   59
## 2011 2012

```

```
## 74 67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 9 6 24 25 31 42 39 32 33 46 56 35 40 49
## 2011 2012
## 60 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 = 55, df = 16, p-value = 4e-06
```

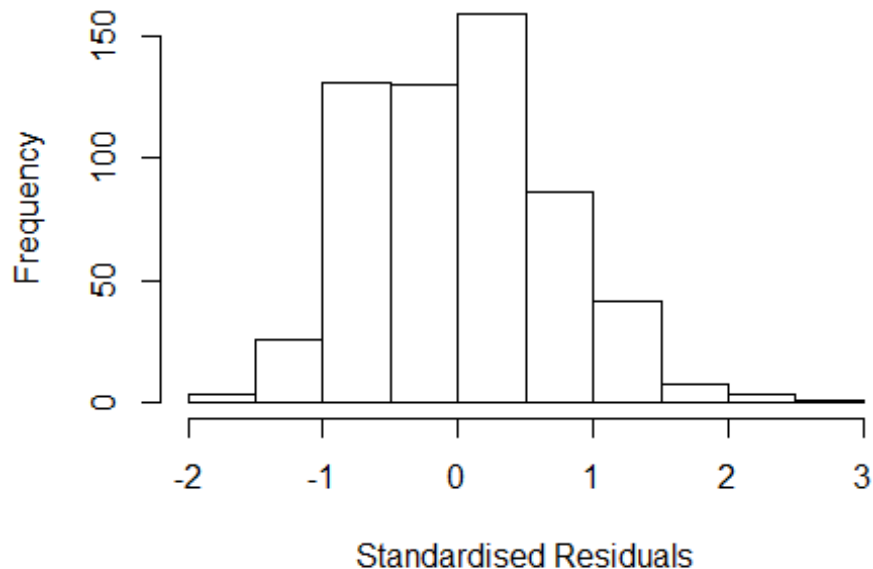


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.9, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.37741436351913"
## [1] "Male first author team size 2018 geometric mean: 2.6684638748106"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.1852256003735"
## [1] "Male last author team size 2018 geometric mean: 2.70548215891423"
##
## 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 1.390 1          1.179
## LastAuthorFemale  1.118 1          1.057
## UniqueAuthors    2.126 4          1.099
## Year              2.379 16         1.027
```

Residuals from first and last author and team size



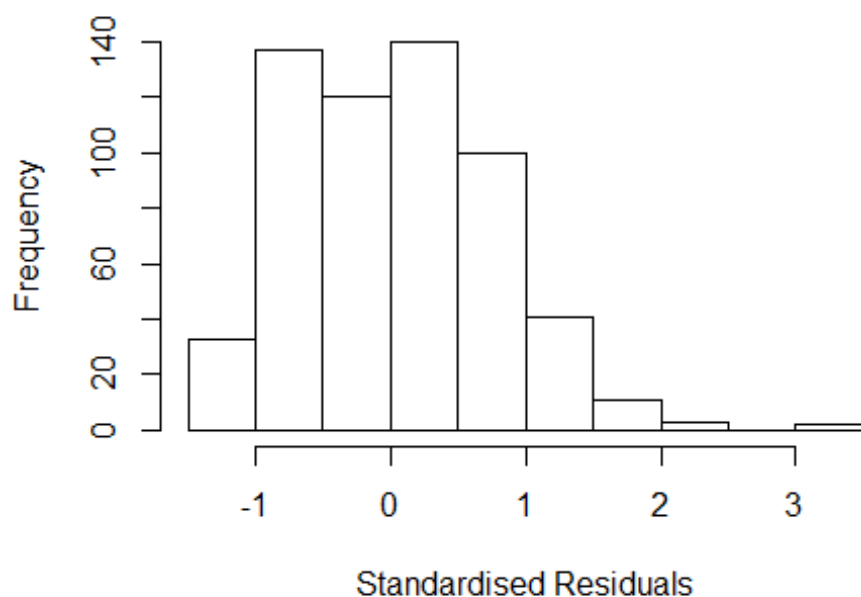
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 599 84930208606 3.94 2001      1909      1      2.63
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6014 -0.5264  0.0184  0.4619  2.6303
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.2685    0.2185   1.23 0.21971
## FirstAuthorFemale1  0.0389    0.0927   0.42 0.67520
## LastAuthorFemale1 -0.1337    0.1249  -1.07 0.28486
## UniqueAuthors2    0.3635    0.0829   4.38 1.4e-05 ***
## UniqueAuthors3    0.7626    0.0782   9.75 < 2e-16 ***
## UniqueAuthors4    0.5661    0.1220   4.64 4.3e-06 ***
## UniqueAuthors5    0.6997    0.1995   3.51 0.00049 ***
## Year1997          0.8147    0.4282   1.90 0.05761 .
## Year1998          0.2333    0.3477   0.67 0.50243
## Year1999          0.5703    0.2939   1.94 0.05279 .
```

```

## Year2000          0.3549      0.2795      1.27  0.20480
## Year2001          0.2787      0.2542      1.10  0.27344
## Year2002          0.3937      0.2662      1.48  0.13969
## Year2003          0.2923      0.2456      1.19  0.23452
## Year2004          0.4461      0.2520      1.77  0.07720 .
## Year2005          0.3842      0.2479      1.55  0.12177
## Year2006          0.0244      0.2360      0.10  0.91754
## Year2007          0.2498      0.2286      1.09  0.27500
## Year2008          0.5628      0.2535      2.22  0.02680 *
## Year2009          0.3689      0.2453      1.50  0.13313
## Year2010          0.4332      0.2388      1.81  0.07021 .
## Year2011          0.5520      0.2265      2.44  0.01511 *
## Year2012          0.6724      0.2357      2.85  0.00449 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.679
## Multiple R-squared:  0.227, Adjusted R-squared:  0.197
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.876  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      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 1.232 1      1.110
## LastAuthorFemale  1.079 1      1.039
## Year              1.304 16      1.008

```

Residuals from first and last author

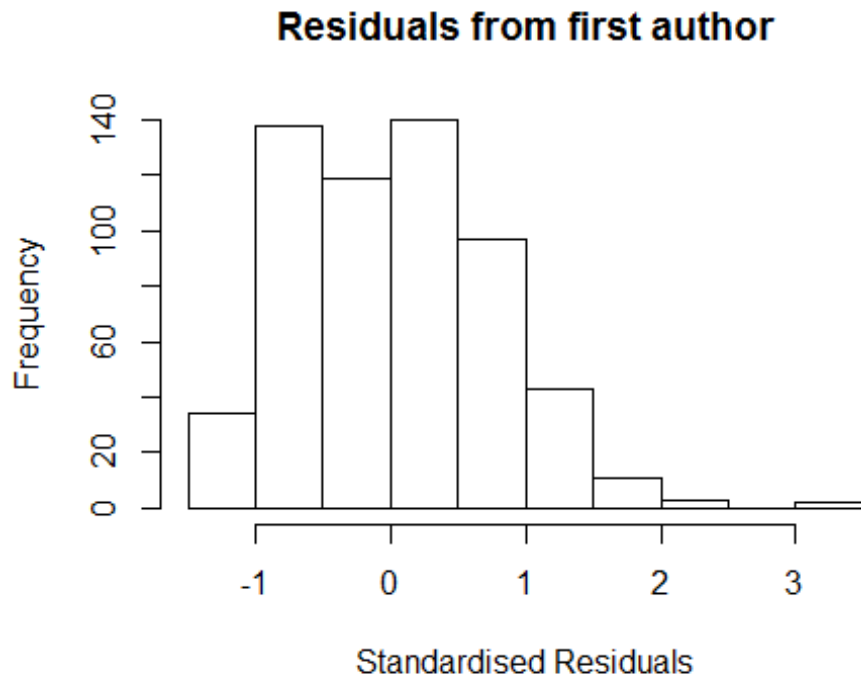


```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 457  0034290459 3.761 2000    1909    1    3.018
## 599  84930208606 3.940 2001    1909    1    3.202
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.44877 -0.59261  0.00264  0.55172  3.20246
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5457    0.2592   2.11  0.0357 *
## FirstAuthorFemale1 0.2059    0.0959   2.15  0.0321 *
## LastAuthorFemale1 -0.0945    0.1397  -0.68  0.4992
## Year1997         0.9031    0.4583   1.97  0.0493 *
## Year1998         0.0568    0.5102   0.11  0.9114
## Year1999         0.5658    0.3261   1.73  0.0833 .
## Year2000         0.1977    0.3250   0.61  0.5433
## Year2001         0.1918    0.2964   0.65  0.5177
## Year2002         0.3614    0.3270   1.10  0.2697
## Year2003         0.3040    0.2815   1.08  0.2806
## Year2004         0.4279    0.2879   1.49  0.1379
```

```

## Year2005          0.4316      0.2979      1.45      0.1479
## Year2006          0.0540      0.2783      0.19      0.8462
## Year2007          0.3764      0.2739      1.37      0.1698
## Year2008          0.5394      0.2895      1.86      0.0629 .
## Year2009          0.3908      0.2902      1.35      0.1786
## Year2010          0.5239      0.2928      1.79      0.0741 .
## Year2011          0.6721      0.2679      2.51      0.0124 *
## Year2012          0.8347      0.2765      3.02      0.0026 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.721
## Multiple R-squared:  0.087, Adjusted R-squared:  0.0581
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 546 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0102 0.8610 0.9380 0.9040 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      1.70e-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 1.21 1      1.100
## Year              1.21 16      1.006

```

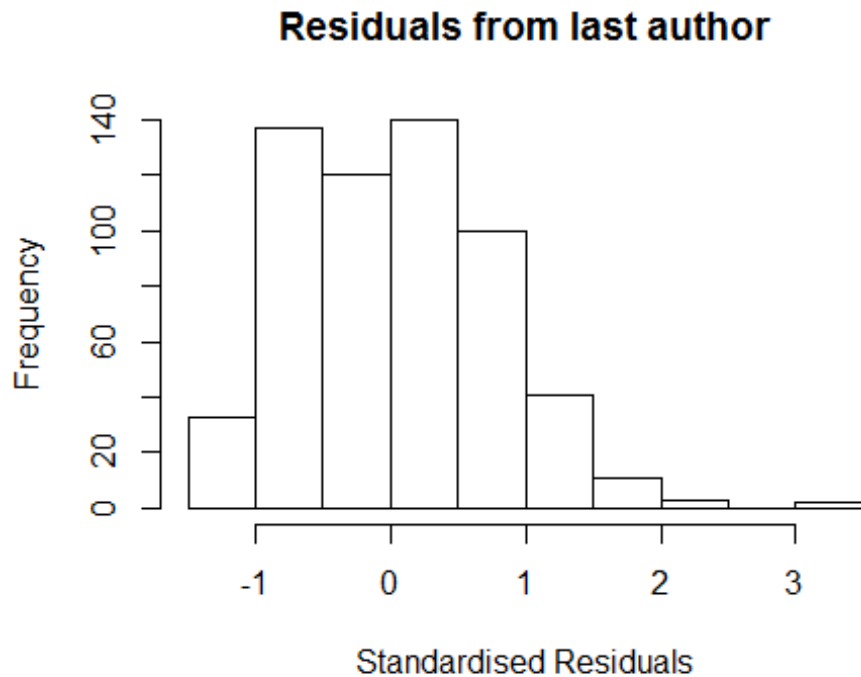



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 457  0034290459 3.761 2000    1909    1    3.018
## 599  84930208606 3.940 2001    1909    1    3.202
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.44928 -0.58579  0.00498  0.55533  3.21310
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5473    0.2597   2.11  0.036 *
## FirstAuthorFemale1 0.1924    0.0971   1.98  0.048 *
## Year1997        0.9020    0.4612   1.96  0.051 .
## Year1998        0.0385    0.5174   0.07  0.941
## Year1999        0.5645    0.3286   1.72  0.086 .
## Year2000        0.1915    0.3264   0.59  0.558
## Year2001        0.1796    0.2973   0.60  0.546
## Year2002        0.3578    0.3287   1.09  0.277
## Year2003        0.2995    0.2825   1.06  0.290
## Year2004        0.4244    0.2883   1.47  0.141
## Year2005        0.4279    0.2980   1.44  0.152
```

```

## Year2006          0.0479      0.2784      0.17      0.863
## Year2007          0.3729      0.2742      1.36      0.174
## Year2008          0.5330      0.2894      1.84      0.066 .
## Year2009          0.3895      0.2913      1.34      0.182
## Year2010          0.5190      0.2941      1.76      0.078 .
## Year2011          0.6657      0.2682      2.48      0.013 *
## Year2012          0.8205      0.2756      2.98      0.003 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.718
## Multiple R-squared:  0.0866, Adjusted R-squared:  0.0593
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 551 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0076 0.8600 0.9380 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      1.70e-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 1.065 1      1.032
## Year      1.065 16      1.002

```



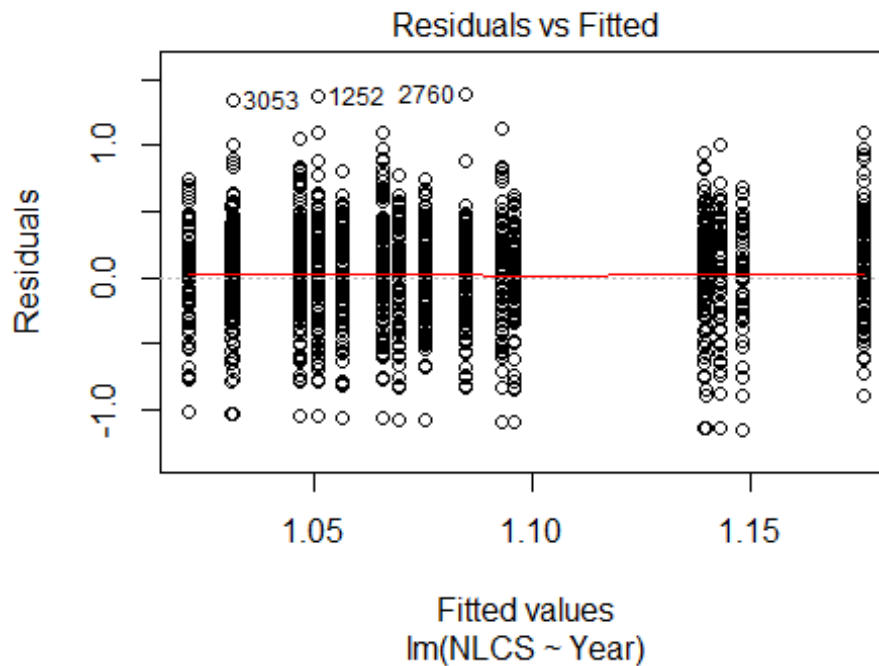
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 457  0034290459 3.761 2000    1909    1    3.018
## 599  84930208606 3.940 2001    1909    1    3.202
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4497 -0.5903  0.0267  0.5437  3.2098
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5776    0.2634   2.19  0.0287 *
## LastAuthorFemale1 -0.0388    0.1325  -0.29  0.7697
## Year1997        0.8721    0.4611   1.89  0.0591 .
## Year1998        0.0533    0.4993   0.11  0.9150
## Year1999        0.5379    0.3290   1.64  0.1026
## Year2000        0.1699    0.3267   0.52  0.6033
## Year2001        0.1526    0.3000   0.51  0.6112
## Year2002        0.3461    0.3350   1.03  0.3019
## Year2003        0.3006    0.2871   1.05  0.2956
## Year2004        0.4015    0.2925   1.37  0.1704
## Year2005        0.4114    0.3015   1.36  0.1729
```

```

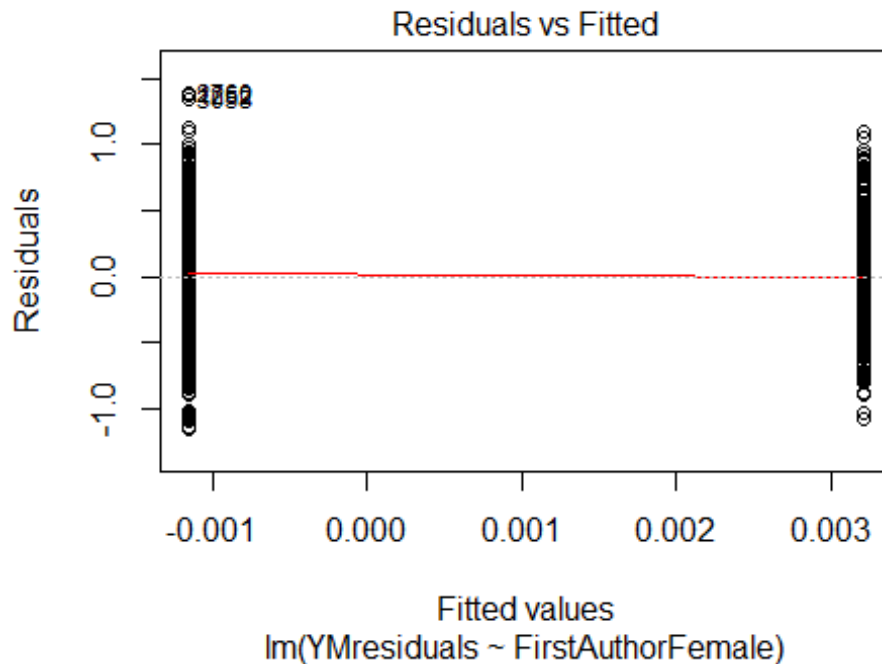
## Year2006          0.0333      0.2819      0.12      0.9061
## Year2007          0.3747      0.2781      1.35      0.1785
## Year2008          0.5178      0.2941      1.76      0.0788 .
## Year2009          0.4210      0.2925      1.44      0.1507
## Year2010          0.5126      0.2969      1.73      0.0848 .
## Year2011          0.6810      0.2716      2.51      0.0124 *
## Year2012          0.8013      0.2802      2.86      0.0044 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.715
## Multiple R-squared:  0.0814, Adjusted R-squared:  0.054
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 549 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0068 0.8540 0.9360 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      1.70e-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: 587"
## [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
## 154 165 176 163 153 153 181 173 165 129 182 212 213 223 254
## 2011 2012
## 237 246
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 93 88 87 62 47 113 109 106 82 118 146 148 139 175

```

```
## 2011 2012
## 143 153
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 87 79 79 54 44 103 102 96 76 105 129 135 132 161
## 2011 2012
## 132 139
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 40, df = 16, p-value = 9e-04
```

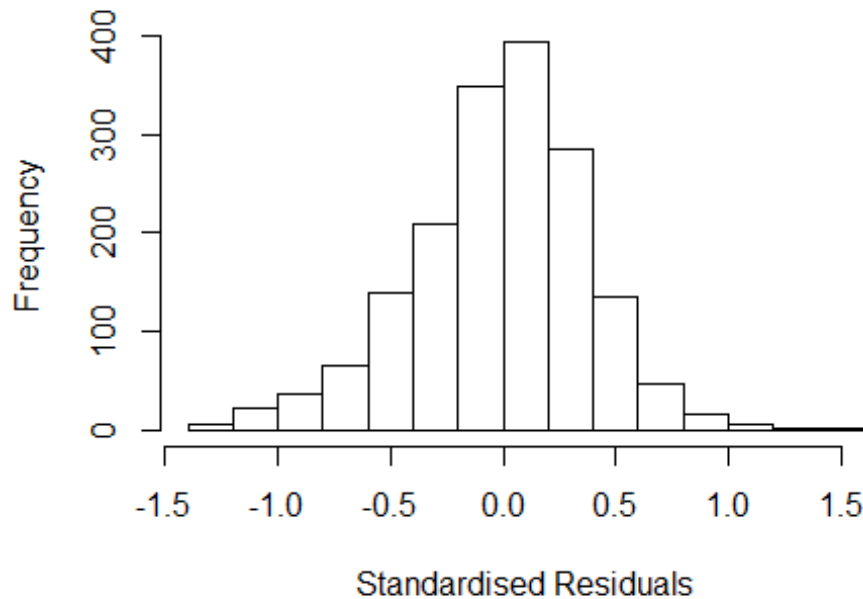


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 7e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.82435374674174"
## [1] "Male first author team size 2018 geometric mean: 3.12500412482468"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77556443359137"
## [1] "Male last author team size 2018 geometric mean: 3.32782907621323"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, 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 1.106 1      1.052
## LastAuthorFemale  1.088 1      1.043
## UniqueAuthors     1.276 4      1.031
## Year              1.369 16     1.010
```

Residuals from first and last author and team size



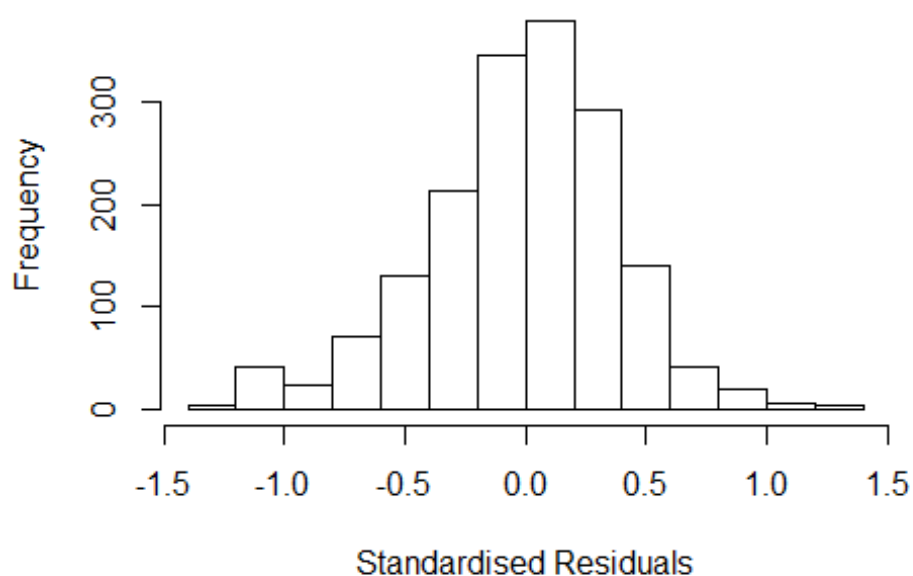
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3262 -0.2388 0.0121 0.2411 1.4943
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1015 0.0684 16.11 < 2e-16 ***
## FirstAuthorFemale1 -0.0236 0.0203 -1.16 0.2455
## LastAuthorFemale1 0.0177 0.0219 0.81 0.4193
## UniqueAuthors2 0.1213 0.0306 3.96 7.8e-05 ***
## UniqueAuthors3 0.1243 0.0310 4.01 6.5e-05 ***
## UniqueAuthors4 0.1451 0.0335 4.33 1.6e-05 ***
## UniqueAuthors5 0.2384 0.0369 6.46 1.4e-10 ***
## Year1997 -0.0795 0.0780 -1.02 0.3080
## Year1998 -0.0137 0.0786 -0.17 0.8618
## Year1999 -0.1320 0.0825 -1.60 0.1095
```

```

## Year2000          -0.0045      0.0851   -0.05   0.9578
## Year2001           0.0299      0.0872    0.34   0.7315
## Year2002          -0.1698      0.0775   -2.19   0.0287 *
## Year2003          -0.1115      0.0756   -1.47   0.1405
## Year2004          -0.1262      0.0762   -1.66   0.0978 .
## Year2005          -0.1603      0.0736   -2.18   0.0295 *
## Year2006          -0.0936      0.0713   -1.31   0.1895
## Year2007          -0.1317      0.0708   -1.86   0.0632 .
## Year2008          -0.1363      0.0708   -1.93   0.0544 .
## Year2009          -0.1934      0.0730   -2.65   0.0081 **
## Year2010          -0.1712      0.0731   -2.34   0.0192 *
## Year2011          -0.0696      0.0752   -0.92   0.3553
## Year2012          -0.1710      0.0738   -2.32   0.0206 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.0486, Adjusted R-squared:  0.0362
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1564 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0415 0.8590 0.9510 0.8920 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          5.84e-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 1.100 1          1.049
## LastAuthorFemale 1.083 1          1.041
## Year 1.095 16          1.003

```


Residuals from first and last author



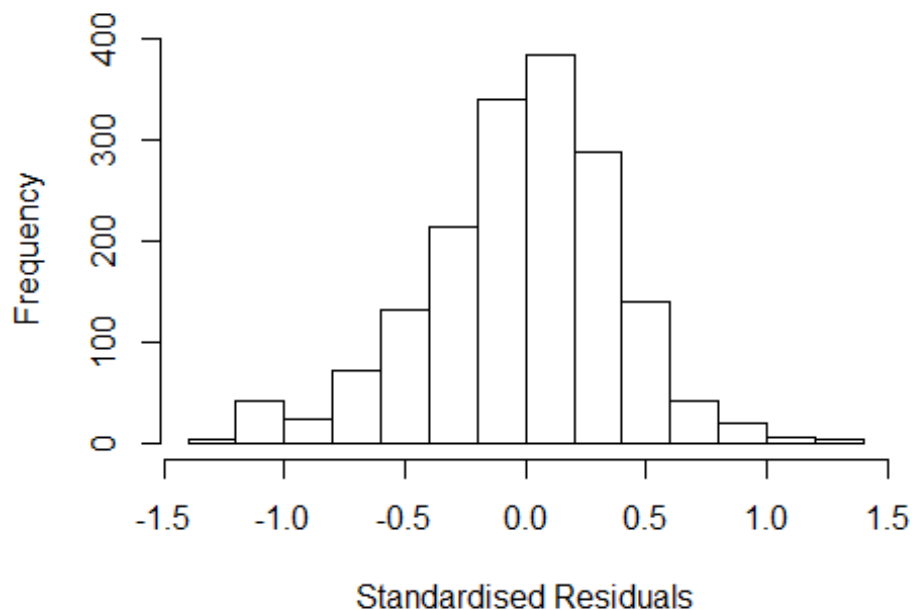
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2379 -0.2480  0.0125  0.2435  1.3901
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18176    0.06651   17.77  <2e-16 ***
## FirstAuthorFemale1 -0.01414    0.02075   -0.68    0.496
## LastAuthorFemale1  0.01591    0.02235    0.71    0.477
## Year1997        -0.07293    0.07874   -0.93    0.354
## Year1998        -0.02301    0.07993   -0.29    0.773
## Year1999        -0.12740    0.08393   -1.52    0.129
## Year2000         0.01028    0.08721    0.12    0.906
## Year2001         0.05615    0.08594    0.65    0.514
## Year2002        -0.14161    0.07927   -1.79    0.074 .
## Year2003        -0.09014    0.07545   -1.19    0.232
## Year2004        -0.09873    0.07806   -1.26    0.206
## Year2005        -0.12386    0.07439   -1.66    0.096 .
```

```

## Year2006          -0.06475    0.07240   -0.89    0.371
## Year2007          -0.10027    0.07145   -1.40    0.161
## Year2008          -0.10090    0.07186   -1.40    0.160
## Year2009          -0.15098    0.07417   -2.04    0.042 *
## Year2010          -0.12083    0.07408   -1.63    0.103
## Year2011          -0.00849    0.07589   -0.11    0.911
## Year2012          -0.11733    0.07518   -1.56    0.119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0186, Adjusted R-squared:  0.00814
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1572 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0989 0.8620 0.9490 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      5.84e-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 1.056 1          1.028
## Year              1.056 16          1.002

```

Residuals from first author



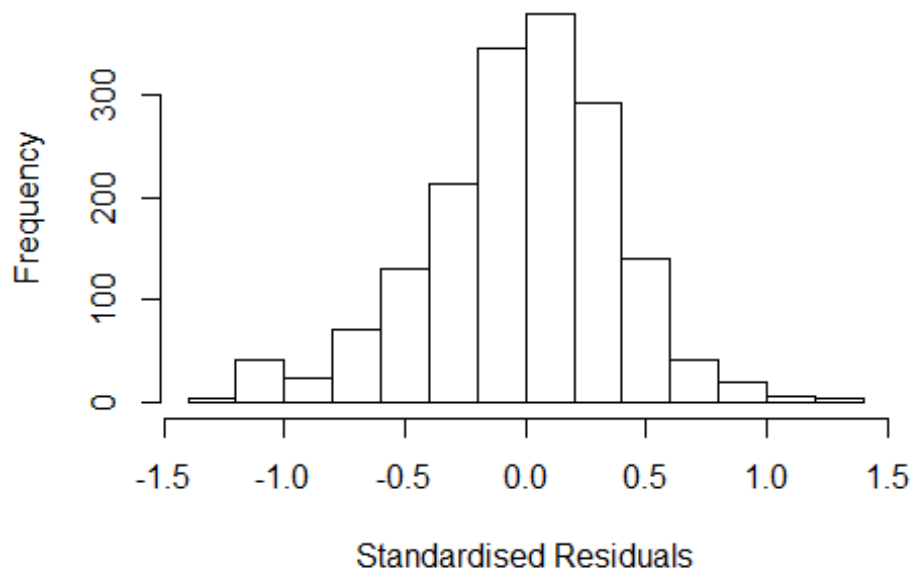
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2395 -0.2497 0.0153 0.2420 1.3879
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18397 0.06618 17.89 <2e-16 ***
## FirstAuthorFemale1 -0.01054 0.02033 -0.52 0.604
## Year1997 -0.07299 0.07853 -0.93 0.353
## Year1998 -0.02322 0.07976 -0.29 0.771
## Year1999 -0.12859 0.08374 -1.54 0.125
## Year2000 0.00961 0.08709 0.11 0.912
## Year2001 0.05555 0.08568 0.65 0.517
## Year2002 -0.14087 0.07911 -1.78 0.075 .
## Year2003 -0.08993 0.07530 -1.19 0.233
## Year2004 -0.09929 0.07785 -1.28 0.202
## Year2005 -0.12417 0.07426 -1.67 0.095 .
## Year2006 -0.06442 0.07224 -0.89 0.373
```

```

## Year2007          -0.09950    0.07131   -1.40    0.163
## Year2008          -0.10083    0.07169   -1.41    0.160
## Year2009          -0.15056    0.07402   -2.03    0.042 *
## Year2010          -0.12011    0.07389   -1.63    0.104
## Year2011          -0.00898    0.07575   -0.12    0.906
## Year2012          -0.11597    0.07505   -1.55    0.122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0183, Adjusted R-squared:  0.00847
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1575 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0996 0.8610 0.9490 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      5.84e-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 1.038 1          1.019
## Year            1.038 16          1.001

```

Residuals from last author



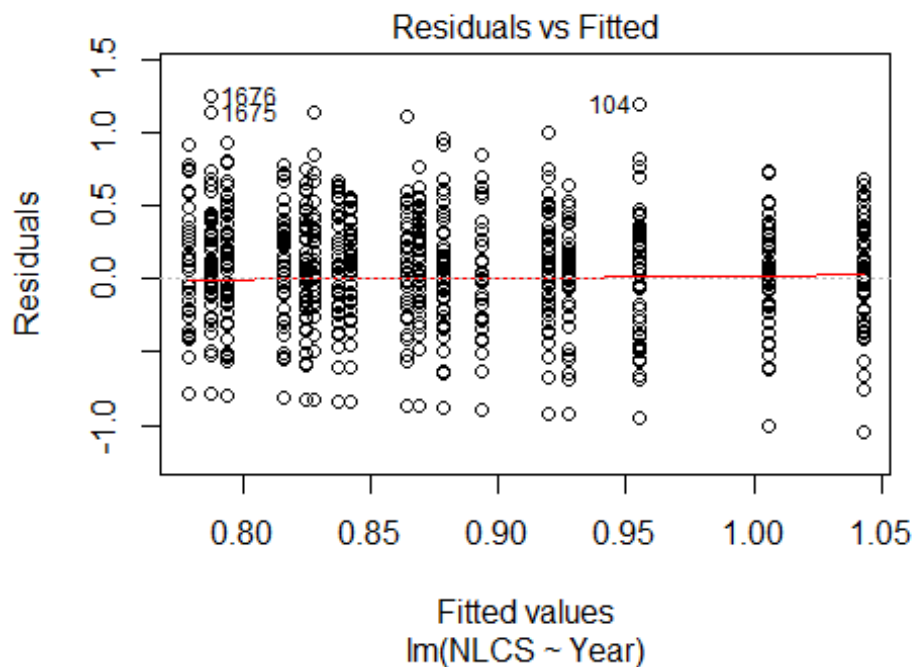
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.233 -0.248  0.015  0.246  1.394
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1792     0.0666   17.72  <2e-16 ***
## LastAuthorFemale1  0.0121     0.0220    0.55   0.581
## Year1997         -0.0716     0.0788   -0.91   0.364
## Year1998         -0.0223     0.0801   -0.28   0.781
## Year1999         -0.1282     0.0840   -1.53   0.127
## Year2000          0.0108     0.0874    0.12   0.902
## Year2001          0.0540     0.0862    0.63   0.531
## Year2002         -0.1412     0.0794   -1.78   0.076 .
## Year2003         -0.0911     0.0755   -1.21   0.228
## Year2004         -0.0994     0.0782   -1.27   0.204
## Year2005         -0.1246     0.0746   -1.67   0.095 .
## Year2006         -0.0655     0.0726   -0.90   0.367
```

```

## Year2007          -0.1009      0.0717   -1.41    0.159
## Year2008          -0.1021      0.0720   -1.42    0.156
## Year2009          -0.1527      0.0743   -2.05    0.040 *
## Year2010          -0.1235      0.0741   -1.67    0.096 .
## Year2011          -0.0101      0.0761   -0.13    0.895
## Year2012          -0.1178      0.0754   -1.56    0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.00837
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1570 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0997 0.8620 0.9490 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      5.84e-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: 1713"
## [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
##   92  131  103  112  109  133   90   85   93   93  106   86   70   98   65
## 2011 2012
##   71   68
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   49   64   56   66   36   56   50   55   49   58   67   55   44   67   45
## 2011 2012

```

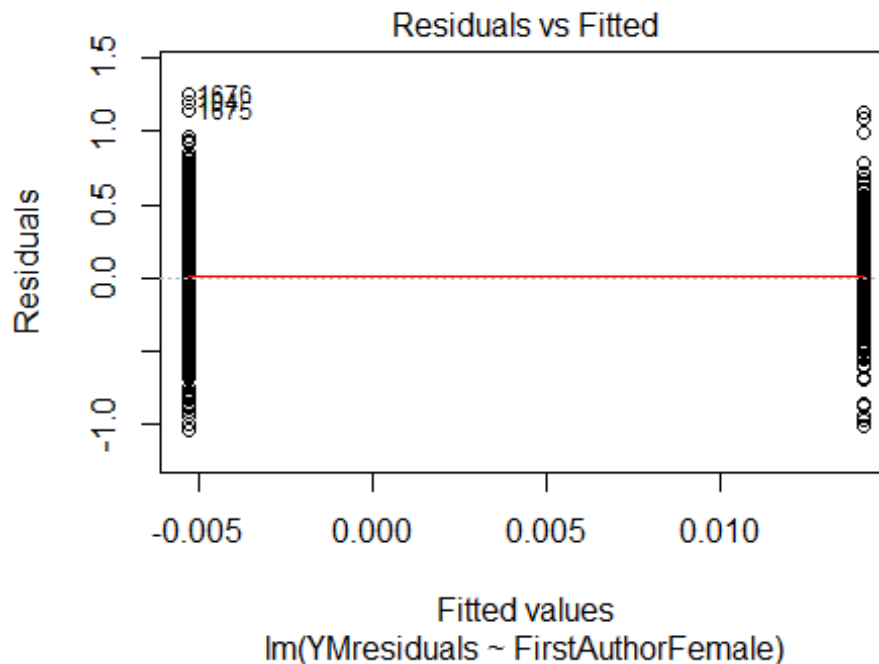
```
## 44 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 48 60 56 63 33 54 43 54 45 54 62 51 41 63 42
## 2011 2012
## 41 40
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: 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 = 4.6, df = 1, p-value = 0.03
## [1] "Female first author team size 2018 geometric mean: 3.13843742524262"
## [1] "Male first author team size 2018 geometric mean: 1.94311853008969"
## 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 = 450, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.69838132956495"
## [1] "Male last author team size 2018 geometric mean: 2.66593769292083"

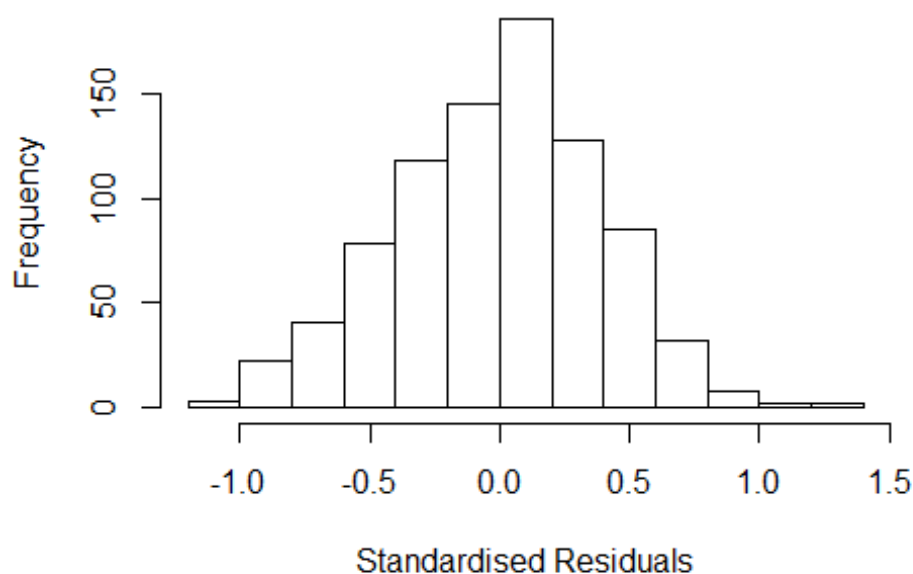
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 84, 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	1.597	1	1.264
LastAuthorFemale	1.588	1	1.260
UniqueAuthors	1.633	4	1.063
Year	1.808	16	1.019

Residuals from first and last author and team size



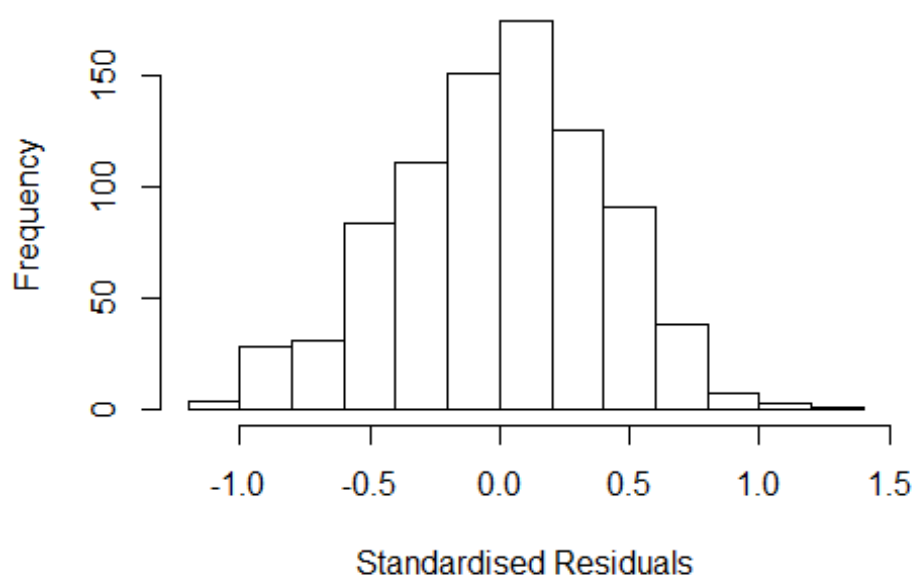
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.197 -0.288 0.024 0.272 1.257
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92449 0.07054 13.11 < 2e-16 ***
## FirstAuthorFemale1 0.07772 0.03627 2.14 0.0324 *
## LastAuthorFemale1 -0.10535 0.03937 -2.68 0.0076 **
## UniqueAuthors2 0.07724 0.03401 2.27 0.0234 *
## UniqueAuthors3 0.18635 0.04530 4.11 4.3e-05 ***
## UniqueAuthors4 0.26924 0.06738 4.00 7.0e-05 ***
## UniqueAuthors5 0.18304 0.06614 2.77 0.0058 **
## Year1997 -0.03194 0.08683 -0.37 0.7131
## Year1998 -0.10786 0.08551 -1.26 0.2076
## Year1999 -0.18260 0.09175 -1.99 0.0469 *
```

```

## Year2000      -0.08736    0.10025   -0.87    0.3838
## Year2001      0.04831    0.09280    0.52    0.6028
## Year2002     -0.18240    0.09695   -1.88    0.0603 .
## Year2003     -0.13973    0.08190   -1.71    0.0884 .
## Year2004     -0.14766    0.08946   -1.65    0.0992 .
## Year2005     -0.13184    0.08752   -1.51    0.1324
## Year2006     -0.13422    0.08314   -1.61    0.1069
## Year2007     -0.00465    0.08529   -0.05    0.9565
## Year2008     -0.07472    0.09746   -0.77    0.4435
## Year2009     -0.22224    0.08741   -2.54    0.0112 *
## Year2010     -0.19297    0.10383   -1.86    0.0635 .
## Year2011     -0.23185    0.09538   -2.43    0.0153 *
## Year2012     -0.13502    0.09784   -1.38    0.1680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0849, Adjusted R-squared:  0.0606
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 780 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.880  0.947  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      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 1.553 1      1.246
## LastAuthorFemale  1.497 1      1.224
## Year              1.210 16      1.006

```

Residuals from first and last author



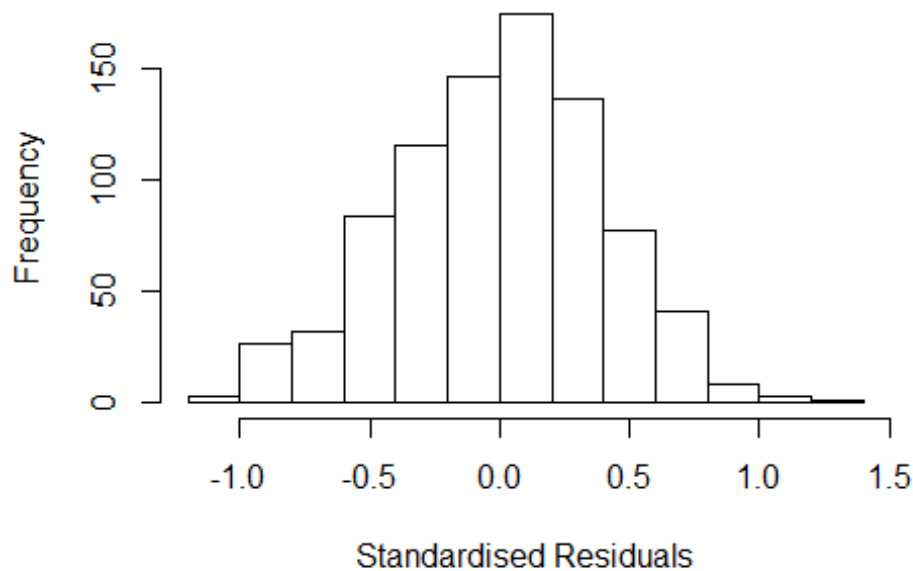
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0580 -0.2799 0.0216 0.2719 1.2492
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9793 0.0721 13.57 <2e-16 ***
## FirstAuthorFemale1 0.0890 0.0363 2.45 0.0145 *
## LastAuthorFemale1 -0.1161 0.0381 -3.05 0.0024 **
## Year1997 -0.0409 0.0881 -0.46 0.6426
## Year1998 -0.1136 0.0880 -1.29 0.1970
## Year1999 -0.1595 0.0923 -1.73 0.0842 .
## Year2000 -0.0602 0.1074 -0.56 0.5753
## Year2001 0.0787 0.0925 0.85 0.3950
## Year2002 -0.1771 0.0986 -1.80 0.0729 .
## Year2003 -0.1346 0.0858 -1.57 0.1170
## Year2004 -0.1488 0.0909 -1.64 0.1022
## Year2005 -0.1241 0.0881 -1.41 0.1593
```

```

## Year2006          -0.1170      0.0873   -1.34   0.1804
## Year2007           0.0293      0.0868    0.34   0.7358
## Year2008          -0.0326      0.0947   -0.34   0.7309
## Year2009          -0.1925      0.0904   -2.13   0.0335 *
## Year2010          -0.1758      0.1104   -1.59   0.1117
## Year2011          -0.1973      0.1008   -1.96   0.0508 .
## Year2012          -0.1076      0.1026   -1.05   0.2948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0459, Adjusted R-squared:  0.0253
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 774 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.330  0.877   0.949   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      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 1.087 1         1.043
## Year              1.087 16         1.003

```

Residuals from first author



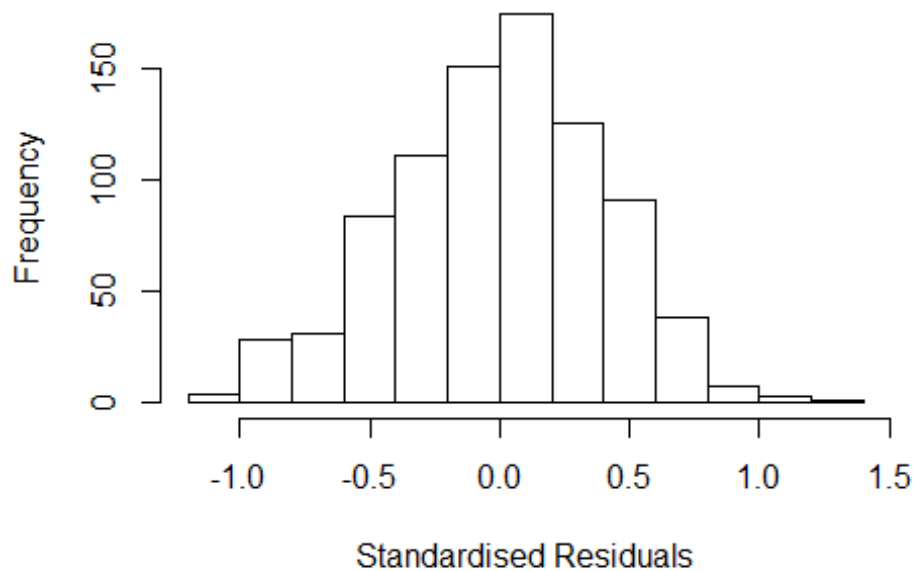
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0522 -0.2915  0.0251  0.2808  1.2556
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9614    0.0690   13.94  <2e-16 ***
## FirstAuthorFemale1 0.0316    0.0307    1.03   0.303
## Year1997      -0.0474    0.0860   -0.55   0.582
## Year1998      -0.1022    0.0853   -1.20   0.231
## Year1999      -0.1615    0.0908   -1.78   0.076 .
## Year2000      -0.0510    0.1065   -0.48   0.632
## Year2001       0.0907    0.0906    1.00   0.317
## Year2002      -0.1703    0.0980   -1.74   0.083 .
## Year2003      -0.1300    0.0847   -1.53   0.126
## Year2004      -0.1529    0.0888   -1.72   0.085 .
## Year2005      -0.1126    0.0851   -1.32   0.186
## Year2006      -0.1105    0.0849   -1.30   0.194
```

```

## Year2007          0.0354      0.0846      0.42      0.676
## Year2008          -0.0264      0.0941     -0.28      0.779
## Year2009          -0.1810      0.0876     -2.07      0.039 *
## Year2010          -0.1559      0.1071     -1.46      0.146
## Year2011          -0.1750      0.0983     -1.78      0.075 .
## Year2012          -0.1060      0.0993     -1.07      0.286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0159
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 784 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.329  0.878  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      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 1.054 1      1.026
## Year              1.054 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
## -1.0654 -0.2920 0.0211 0.2723 1.2269
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9839 0.0716 13.74 <2e-16 ***
## LastAuthorFemale1 -0.0671 0.0322 -2.08 0.038 *
## Year1997 -0.0446 0.0882 -0.51 0.613
## Year1998 -0.1019 0.0873 -1.17 0.244
## Year1999 -0.1512 0.0924 -1.64 0.102
## Year2000 -0.0550 0.1070 -0.51 0.607
## Year2001 0.0815 0.0917 0.89 0.375
## Year2002 -0.1710 0.0991 -1.73 0.085 .
## Year2003 -0.1285 0.0853 -1.51 0.132
## Year2004 -0.1419 0.0909 -1.56 0.119
## Year2005 -0.1138 0.0875 -1.30 0.194
## Year2006 -0.1046 0.0865 -1.21 0.227
```

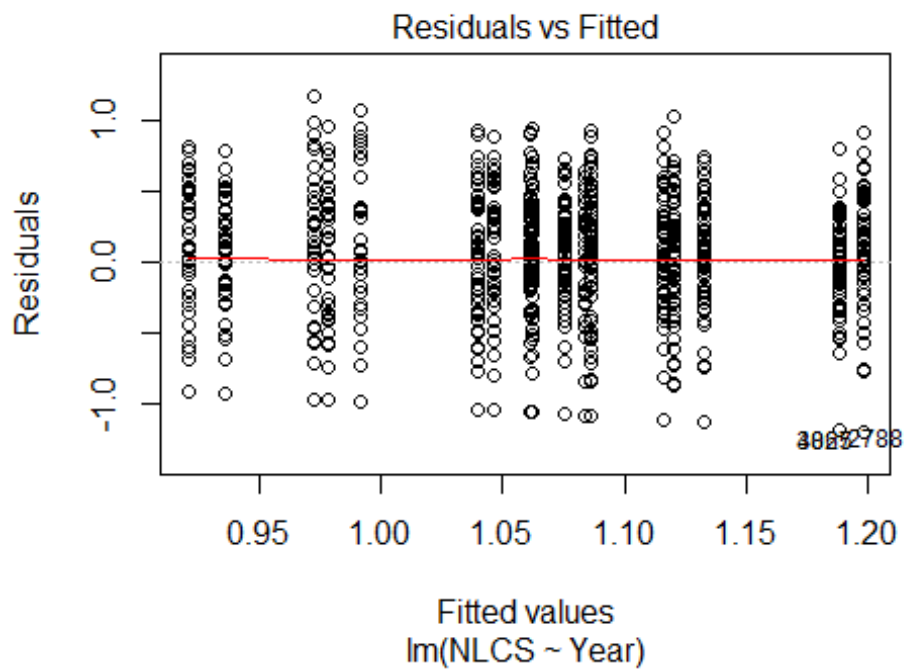
```

## Year2007          0.0445      0.0856      0.52      0.603
## Year2008          -0.0219     0.0947     -0.23     0.817
## Year2009          -0.1748     0.0891     -1.96     0.050 .
## Year2010          -0.1503     0.1085     -1.39     0.166
## Year2011          -0.1782     0.0997     -1.79     0.074 .
## Year2012          -0.1013     0.1018     -0.99     0.320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0389, Adjusted R-squared:  0.0192
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 776 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.350  0.872  0.945  0.909  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      1.18e-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: 850"
## [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
## 126 104 141 97 114 136 120 102 137 108 144 120 133 110 130
## 2011 2012
## 141 169
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 36 64 41 47 54 46 40 61 53 67 56 58 55 64
## 2011 2012

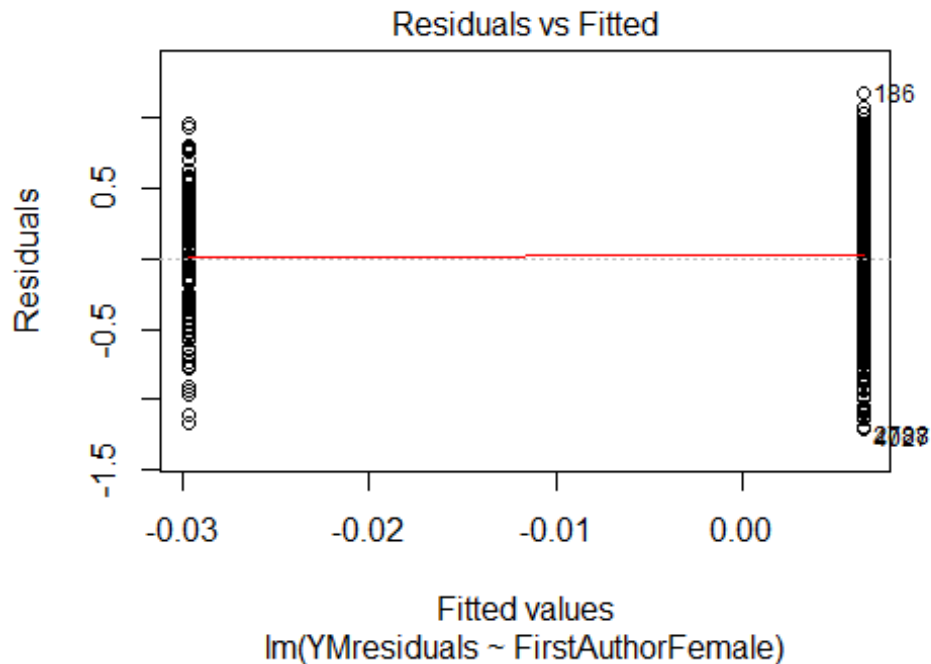
```



```
## 64 72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 31 55 34 43 44 45 37 54 46 53 44 47 49 56
## 2011 2012
## 55 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 = 36, df = 16, p-value = 0.003
```



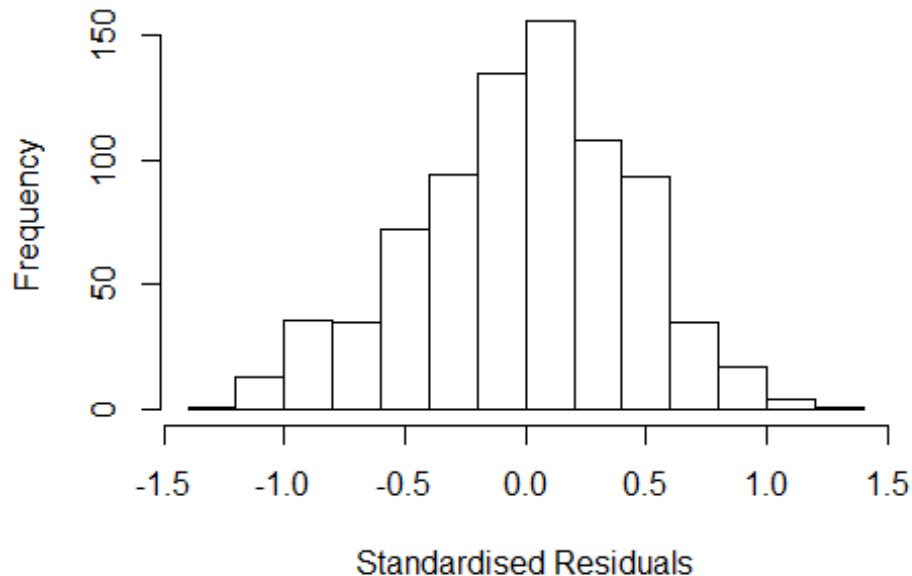
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.5, df = 1, p-value = 0.06
## [1] "Female first author team size 2018 geometric mean: 2.85357080900276"
## [1] "Male first author team size 2018 geometric mean: 2.71334014387658"
## 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 = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.8612097182042"
## [1] "Male last author team size 2018 geometric mean: 2.82133251042069"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 78, 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	1.206	1	1.098
LastAuthorFemale	1.174	1	1.083
UniqueAuthors	1.666	4	1.066
Year	1.809	16	1.019

Residuals from first and last author and team size



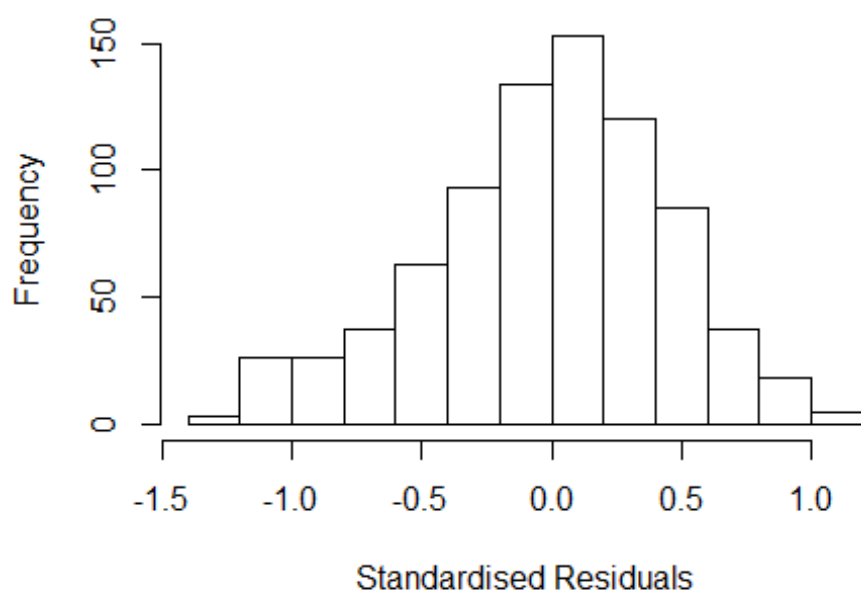
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2199 -0.2969 0.0118 0.2968 1.2821
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86693 0.12744 6.80 2.1e-11 ***
## FirstAuthorFemale1 -0.09936 0.04031 -2.46 0.0139 *
## LastAuthorFemale1 -0.00996 0.04963 -0.20 0.8410
## UniqueAuthors2 0.21057 0.04744 4.44 1.0e-05 ***
## UniqueAuthors3 0.28016 0.05210 5.38 1.0e-07 ***
## UniqueAuthors4 0.27773 0.05661 4.91 1.1e-06 ***
## UniqueAuthors5 0.20710 0.06978 2.97 0.0031 **
## Year1997 0.01117 0.16530 0.07 0.9461
## Year1998 0.13564 0.13588 1.00 0.3185
## Year1999 -0.03293 0.15300 -0.22 0.8296
```

```

## Year2000      -0.12985    0.15012   -0.86    0.3873
## Year2001      0.03779    0.14407    0.26    0.7932
## Year2002      0.01239    0.14708    0.08    0.9329
## Year2003      0.04165    0.14596    0.29    0.7754
## Year2004      0.10930    0.14333    0.76    0.4460
## Year2005      0.05181    0.14153    0.37    0.7144
## Year2006      0.08131    0.13768    0.59    0.5550
## Year2007      0.19043    0.14180    1.34    0.1797
## Year2008      0.19560    0.13197    1.48    0.1387
## Year2009      0.07278    0.13479    0.54    0.5894
## Year2010      0.06192    0.13597    0.46    0.6490
## Year2011     -0.01698    0.13574   -0.13    0.9005
## Year2012      0.14571    0.13426    1.09    0.2781
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0924, Adjusted R-squared:  0.0667
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 722 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.368  0.869  0.947  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      1.25e-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 1.170 1      1.082
## LastAuthorFemale  1.157 1      1.076
## Year              1.198 16      1.006

```

Residuals from first and last author



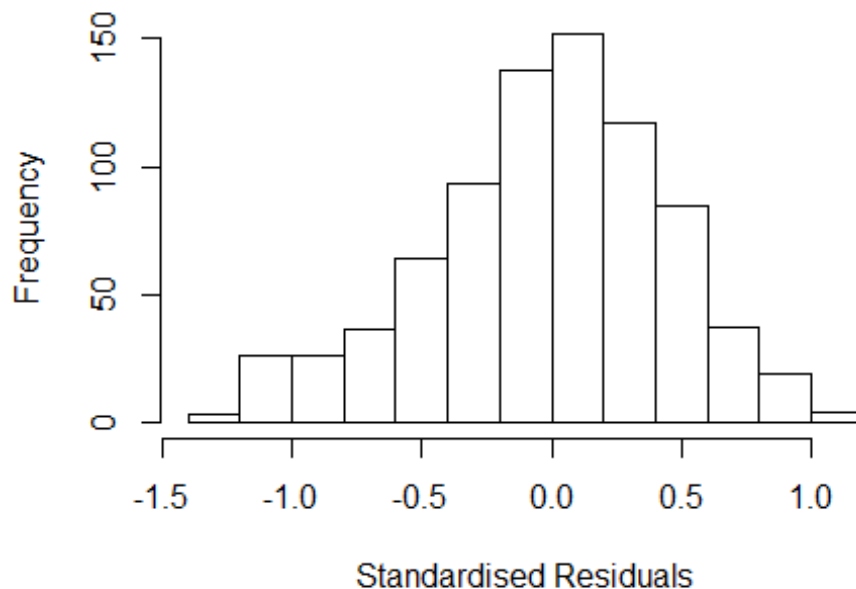
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2252 -0.2888  0.0221  0.2955  1.1795
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96955    0.12828   7.56 1.1e-13 ***
## FirstAuthorFemale1 -0.06979    0.04021  -1.74  0.083 .
## LastAuthorFemale1 -0.01498    0.04802  -0.31  0.755
## Year1997        -0.01341    0.17657  -0.08  0.939
## Year1998         0.15251    0.14154   1.08  0.282
## Year1999        -0.00151    0.15716  -0.01  0.992
## Year2000        -0.06370    0.15498  -0.41  0.681
## Year2001         0.11333    0.14844   0.76  0.445
## Year2002         0.06043    0.14990   0.40  0.687
## Year2003         0.15979    0.14466   1.10  0.270
## Year2004         0.18229    0.15109   1.21  0.228
## Year2005         0.12480    0.14413   0.87  0.387
```

```

## Year2006          0.16426      0.13948      1.18      0.239
## Year2007          0.23406      0.14460      1.62      0.106
## Year2008          0.25569      0.13487      1.90      0.058 .
## Year2009          0.13931      0.13878      1.00      0.316
## Year2010          0.12475      0.13809      0.90      0.367
## Year2011          0.04386      0.14116      0.31      0.756
## Year2012          0.23539      0.13687      1.72      0.086 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.0385, Adjusted R-squared:  0.0164
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 717 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.420  0.858  0.947   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      1.25e-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 1.106 1          1.051
## Year              1.106 16          1.003

```

Residuals from first author



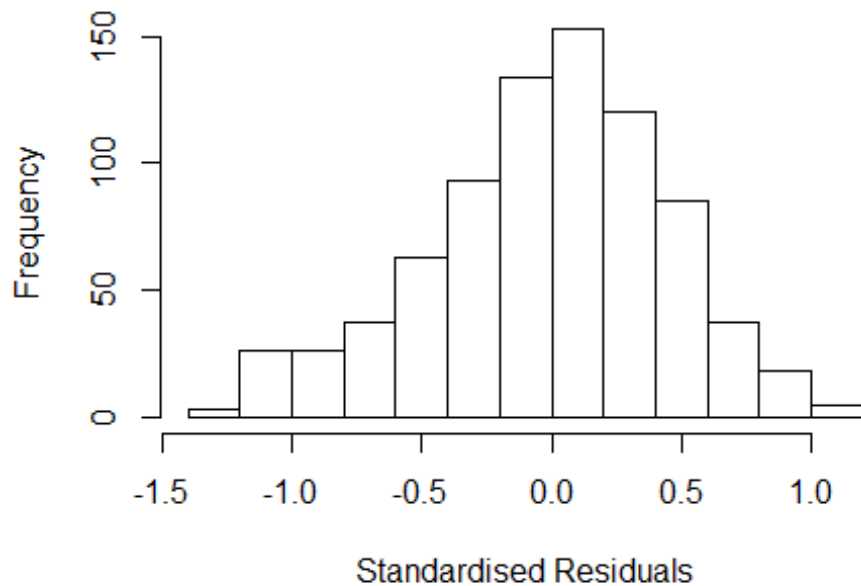
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2239 -0.2878 0.0206 0.2970 1.1810
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.968015 0.128339 7.54 1.3e-13 ***
## FirstAuthorFemale1 -0.073374 0.039151 -1.87 0.061 .
## Year1997 -0.012741 0.176848 -0.07 0.943
## Year1998 0.152167 0.141626 1.07 0.283
## Year1999 0.000333 0.157163 0.00 0.998
## Year2000 -0.064010 0.155119 -0.41 0.680
## Year2001 0.113977 0.148704 0.77 0.444
## Year2002 0.060403 0.149924 0.40 0.687
## Year2003 0.161633 0.144747 1.12 0.264
## Year2004 0.182843 0.151319 1.21 0.227
## Year2005 0.124429 0.144179 0.86 0.388
## Year2006 0.164598 0.139660 1.18 0.239
```

```

## Year2007          0.234915    0.144673    1.62    0.105
## Year2008          0.255932    0.135029    1.90    0.058 .
## Year2009          0.140898    0.138834    1.01    0.310
## Year2010          0.125900    0.138188    0.91    0.363
## Year2011          0.043125    0.141200    0.31    0.760
## Year2012          0.235980    0.136977    1.72    0.085 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0384, Adjusted R-squared:  0.0175
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.420  0.858  0.946  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      1.25e-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 1.093 1          1.045
## Year            1.093 16          1.003

```


Residuals from last author



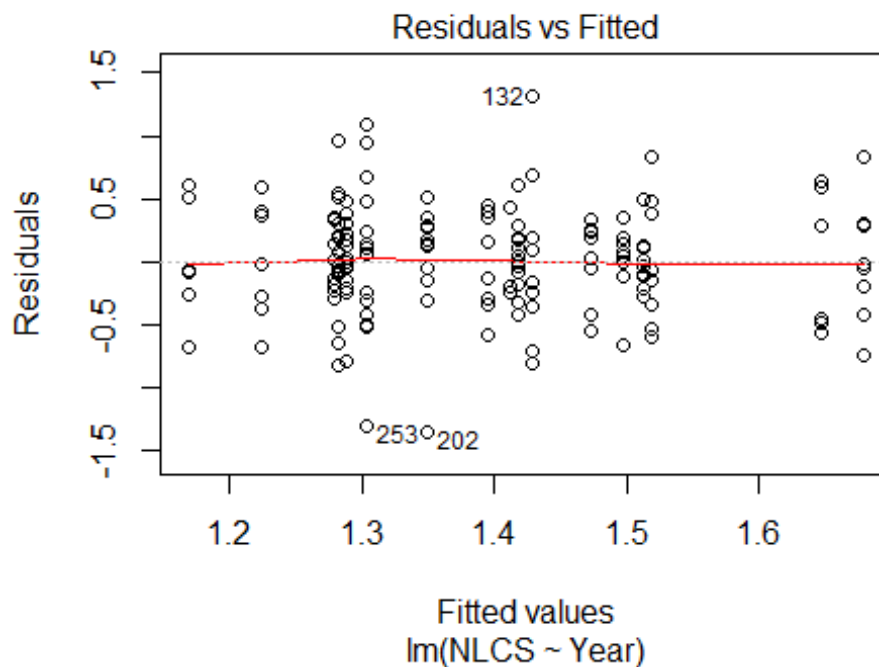
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2153 -0.3028 0.0129 0.2979 1.1866
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96243 0.12856 7.49 1.9e-13 ***
## LastAuthorFemale1 -0.03877 0.04733 -0.82 0.413
## Year1997 -0.00508 0.17681 -0.03 0.977
## Year1998 0.15622 0.14203 1.10 0.272
## Year1999 -0.00429 0.15685 -0.03 0.978
## Year2000 -0.06621 0.15580 -0.42 0.671
## Year2001 0.11441 0.14872 0.77 0.442
## Year2002 0.05374 0.15064 0.36 0.721
## Year2003 0.15255 0.14535 1.05 0.294
## Year2004 0.18032 0.15183 1.19 0.235
## Year2005 0.12575 0.14481 0.87 0.385
## Year2006 0.15242 0.14001 1.09 0.277
```

```

## Year2007      0.23198      0.14498      1.60      0.110
## Year2008      0.25286      0.13561      1.86      0.063 .
## Year2009      0.12414      0.13924      0.89      0.373
## Year2010      0.12008      0.13891      0.86      0.388
## Year2011      0.04505      0.14166      0.32      0.751
## Year2012      0.22779      0.13779      1.65      0.099 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0146
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 732 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.430  0.863  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      1.25e-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 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
##   15   12   11   16   16   17   18    7   11   11   11   20   27   22   25
## 2011 2012
##   18   25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    8    8    9    6    8    9    3    8    8    6   11   15   13   15
## 2011 2012

```

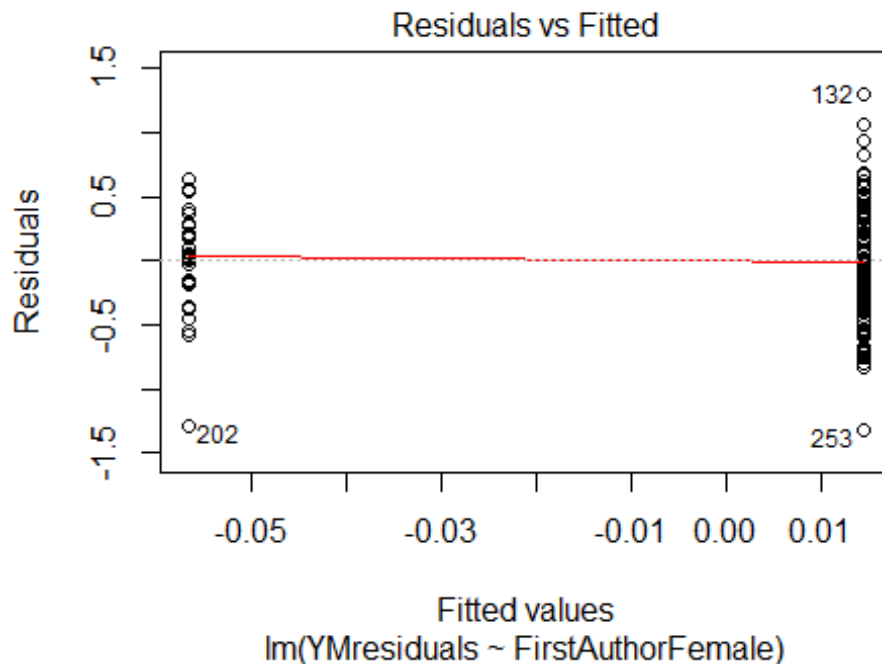
```
##      7    15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      7      7      8      8      5      6      9      3      8      8      5      11      14      13      13
## 2011 2012
##      7      13
## [1] "Heteroscedasticity checks, confirming that there 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 = 0.4, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 4.02469323417112"
## [1] "Male first author team size 2018 geometric mean: 4.11694268393347"
## 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 = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.20086866888094"
## [1] "Male last author team size 2018 geometric mean: 4.52916915378233"

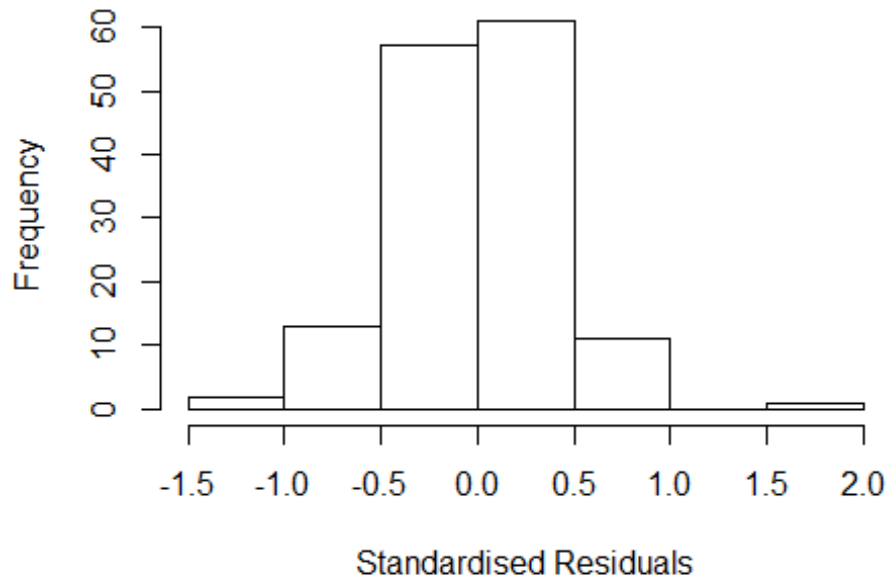
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21, p-value = 0.4
## 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	1.308	1	1.144
LastAuthorFemale	2.181	1	1.477
UniqueAuthors	18.979	4	1.445
Year	29.365	16	1.111

Residuals from first and last author and team size



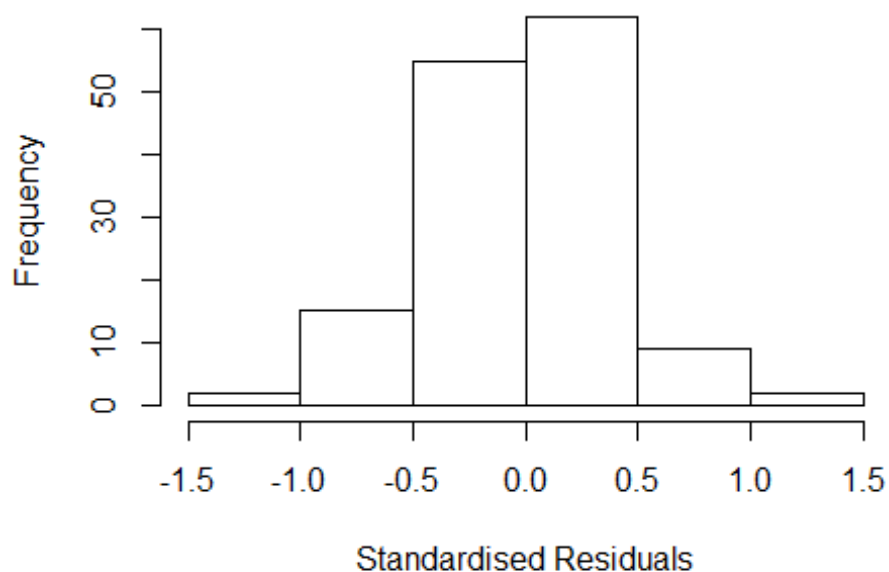
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4189 -0.2280 0.0015 0.2703 1.5889
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29229 0.11825 10.93 <2e-16 ***
## FirstAuthorFemale1 -0.08118 0.07782 -1.04 0.299
## LastAuthorFemale1 -0.03931 0.09918 -0.40 0.693
## UniqueAuthors2 0.23005 0.12105 1.90 0.060 .
## UniqueAuthors3 0.26148 0.12607 2.07 0.040 *
## UniqueAuthors4 0.25512 0.14809 1.72 0.087 .
## UniqueAuthors5 0.17140 0.12204 1.40 0.163
## Year1997 -0.02633 0.18413 -0.14 0.887
## Year1998 0.30400 0.18977 1.60 0.112
## Year1999 0.05972 0.13190 0.45 0.651
```

```

## Year2000          0.34419    0.24598    1.40    0.164
## Year2001          0.06543    0.23653    0.28    0.783
## Year2002         -0.14018    0.26744   -0.52    0.601
## Year2003         -0.14500    0.25430   -0.57    0.570
## Year2004          0.01989    0.15183    0.13    0.896
## Year2005         -0.16542    0.13436   -1.23    0.221
## Year2006         -0.18659    0.29821   -0.63    0.533
## Year2007          0.00533    0.16642    0.03    0.975
## Year2008         -0.10344    0.25134   -0.41    0.681
## Year2009         -0.05411    0.13431   -0.40    0.688
## Year2010         -0.25607    0.17051   -1.50    0.136
## Year2011         -0.23910    0.19956   -1.20    0.233
## Year2012         -0.11253    0.15570   -0.72    0.471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.168, Adjusted R-squared:  0.0176
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 127 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0781 0.8740 0.9570 0.9010 0.9820 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      6.90e-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 1.319 1      1.149
## LastAuthorFemale  1.803 1      1.343
## Year              2.351 16      1.027

```

Residuals from first and last author



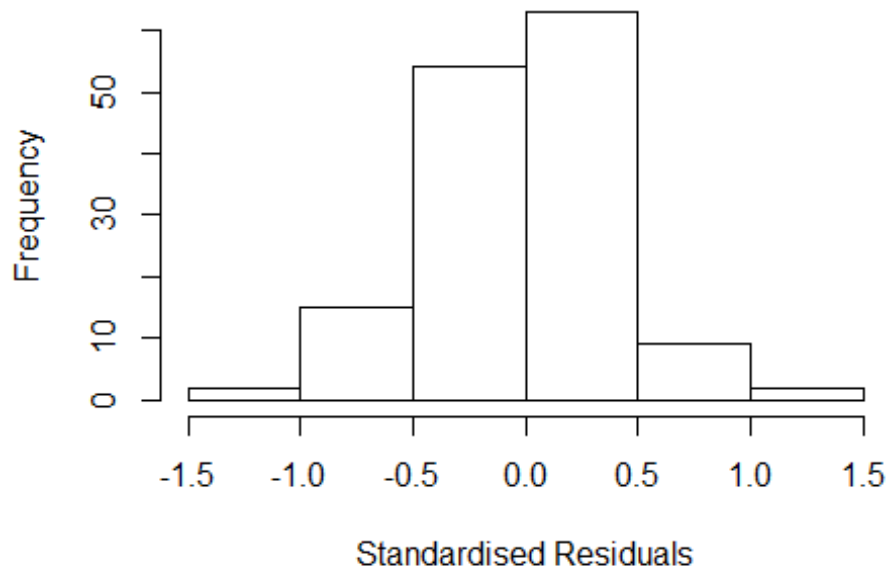
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41417 -0.26821 0.00709 0.28109 1.44419
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4421 0.1139 12.66 <2e-16 ***
## FirstAuthorFemale1 -0.0525 0.0803 -0.65 0.51
## LastAuthorFemale1 -0.0073 0.0902 -0.08 0.94
## Year1997 -0.0686 0.2033 -0.34 0.74
## Year1998 0.2346 0.2097 1.12 0.27
## Year1999 0.0843 0.1379 0.61 0.54
## Year2000 0.3511 0.3240 1.08 0.28
## Year2001 -0.0381 0.2336 -0.16 0.87
## Year2002 -0.1453 0.2518 -0.58 0.57
## Year2003 -0.0407 0.2249 -0.18 0.86
## Year2004 0.0934 0.1520 0.61 0.54
## Year2005 -0.1509 0.1450 -1.04 0.30
```

```

## Year2006          -0.1900      0.2792   -0.68      0.50
## Year2007           0.0319      0.1491    0.21      0.83
## Year2008          -0.0764      0.2475   -0.31      0.76
## Year2009          -0.0192      0.1366   -0.14      0.89
## Year2010          -0.2446      0.1639   -1.49      0.14
## Year2011          -0.2117      0.2232   -0.95      0.34
## Year2012          -0.1077      0.1468   -0.73      0.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.112, Adjusted R-squared:  -0.0152
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 129 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.882  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      6.90e-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"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.362 1      1.167
## Year              1.362 16      1.010

```


Residuals from first author



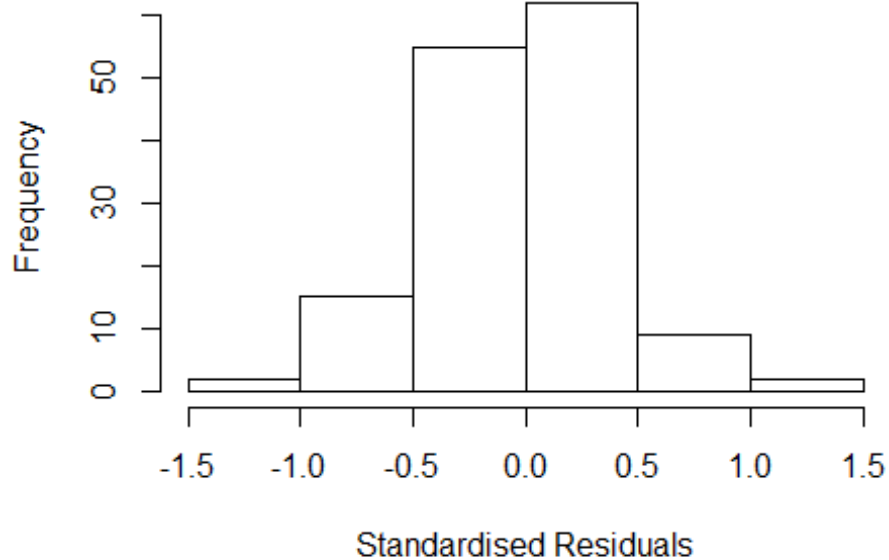
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42127 -0.27431 0.00561 0.27724 1.45085
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4415 0.1144 12.60 <2e-16 ***
## FirstAuthorFemale1 -0.0529 0.0814 -0.65 0.52
## Year1997 -0.0692 0.2047 -0.34 0.74
## Year1998 0.2349 0.2077 1.13 0.26
## Year1999 0.0837 0.1386 0.60 0.55
## Year2000 0.3529 0.3278 1.08 0.28
## Year2001 -0.0394 0.2360 -0.17 0.87
## Year2002 -0.1513 0.2512 -0.60 0.55
## Year2003 -0.0425 0.2246 -0.19 0.85
## Year2004 0.0948 0.1524 0.62 0.54
## Year2005 -0.1521 0.1448 -1.05 0.30
## Year2006 -0.1885 0.2813 -0.67 0.50
```

```

## Year2007          0.0327      0.1460      0.22      0.82
## Year2008         -0.0795      0.2343     -0.34      0.73
## Year2009         -0.0203      0.1371     -0.15      0.88
## Year2010         -0.2447      0.1643     -1.49      0.14
## Year2011         -0.2108      0.2246     -0.94      0.35
## Year2012         -0.1088      0.1455     -0.75      0.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.113, Adjusted R-squared:  -0.00551
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 130 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.174  0.877   0.950   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      6.90e-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 1.861 1      1.364
## Year              1.861 16      1.020

```

Residuals from last author



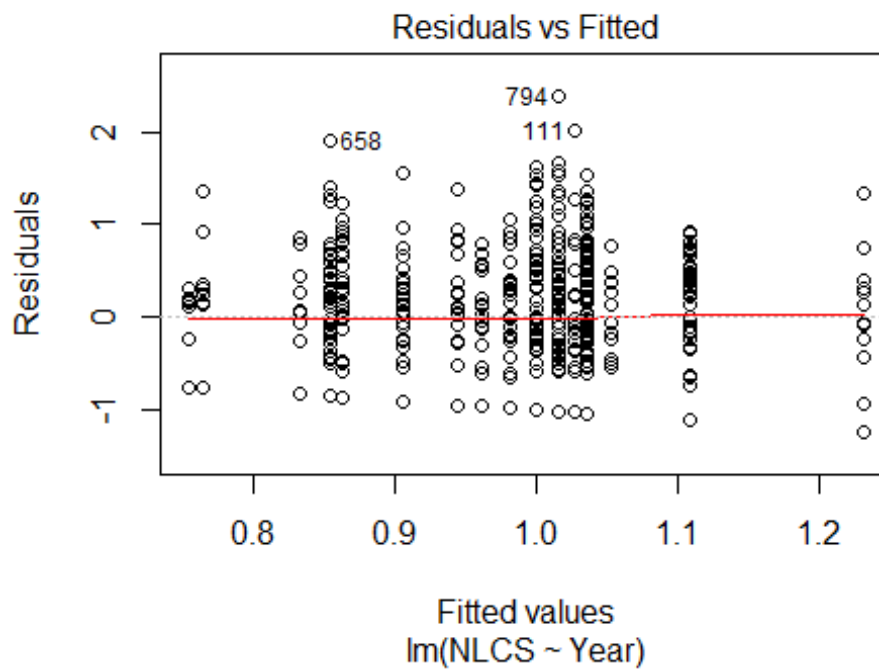
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4466 -0.2643 0.0171 0.2872 1.4395
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4359 0.1173 12.24 <2e-16 ***
## LastAuthorFemale1 -0.0147 0.0935 -0.16 0.88
## Year1997 -0.0612 0.2047 -0.30 0.77
## Year1998 0.2318 0.2111 1.10 0.27
## Year1999 0.0849 0.1426 0.60 0.55
## Year2000 0.3456 0.3130 1.10 0.27
## Year2001 -0.0437 0.2341 -0.19 0.85
## Year2002 -0.1344 0.2543 -0.53 0.60
## Year2003 -0.0322 0.2292 -0.14 0.89
## Year2004 0.0783 0.1494 0.52 0.60
## Year2005 -0.1564 0.1482 -1.06 0.29
## Year2006 -0.1951 0.2710 -0.72 0.47
```

```

## Year2007          0.0254      0.1509      0.17      0.87
## Year2008          -0.0890      0.2483     -0.36      0.72
## Year2009          -0.0191      0.1407     -0.14      0.89
## Year2010          -0.2545      0.1651     -1.54      0.13
## Year2011          -0.2056      0.2244     -0.92      0.36
## Year2012          -0.1211      0.1484     -0.82      0.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.109, Adjusted R-squared:  -0.0104
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.203  0.891  0.951  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      6.90e-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: 145"
## [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
##   20   11   18   16   23   25   30   26   30   35   54   77  101   79  101
## 2011 2012
##  128  110
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14    7   13   13   18   23   25   22   20   28   46   56   78   54   74
## 2011 2012

```

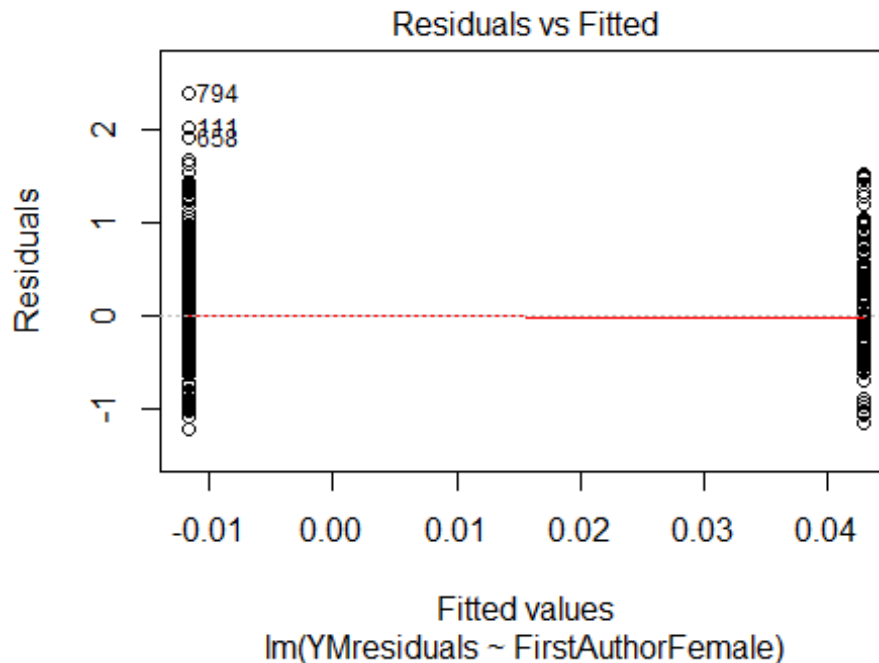
```
## 96 80
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 7 13 13 18 22 22 17 18 27 44 50 71 51 58
## 2011 2012
## 85 69
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.001, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 1.55997934086402"
## [1] "Male first author team size 2018 geometric mean: 1.56310939551105"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 1
```

```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.6424824339403"
## [1] "Male last author team size 2018 geometric mean: 1.53509408843745"

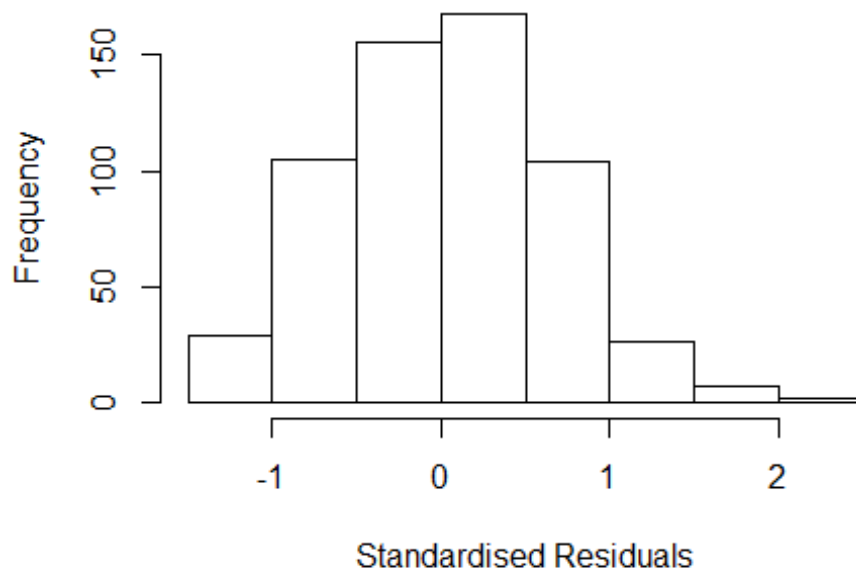
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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	2.089	1	1.445
LastAuthorFemale	2.199	1	1.483
UniqueAuthors	1.604	4	1.061
Year	1.712	16	1.017

Residuals from first and last author and team size



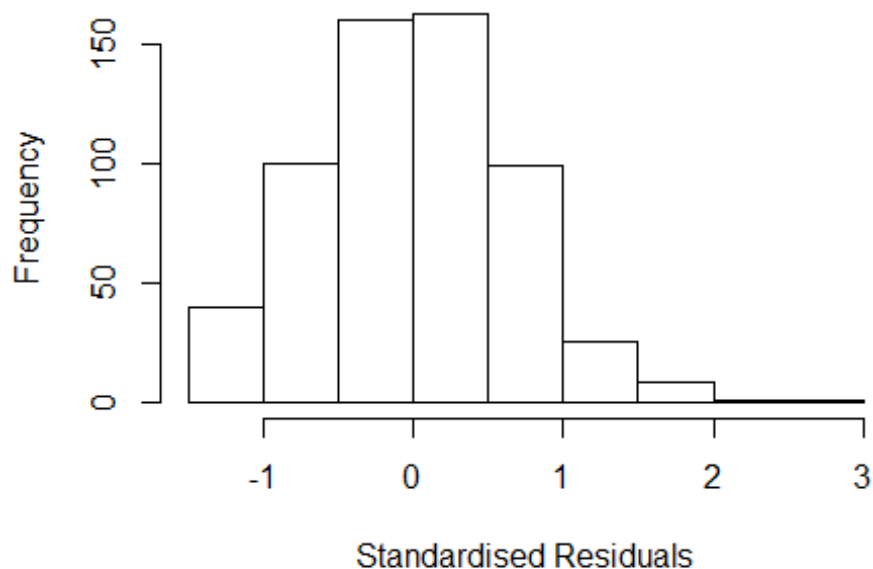
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1844 -0.4655 0.0135 0.4481 2.4501
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.76427 0.21080 3.63 0.00031 ***
## FirstAuthorFemale1 0.00455 0.09697 0.05 0.96260
## LastAuthorFemale1 0.08888 0.10067 0.88 0.37768
## UniqueAuthors2 0.12292 0.06650 1.85 0.06506 .
## UniqueAuthors3 0.17590 0.12441 1.41 0.15795
## UniqueAuthors4 0.29442 0.12929 2.28 0.02314 *
## UniqueAuthors5 0.44537 0.27791 1.60 0.10957
## Year1997 -0.07148 0.24750 -0.29 0.77285
## Year1998 0.23892 0.24463 0.98 0.32916
## Year1999 0.39977 0.27010 1.48 0.13940
```

```

## Year2000      -0.08422    0.27697   -0.30  0.76118
## Year2001      0.05980    0.24456    0.24  0.80690
## Year2002      0.18007    0.25951    0.69  0.48803
## Year2003      0.03544    0.26534    0.13  0.89380
## Year2004      0.06222    0.24073    0.26  0.79613
## Year2005      0.06397    0.24152    0.26  0.79120
## Year2006     -0.01516    0.23572   -0.06  0.94876
## Year2007      0.29060    0.22953    1.27  0.20600
## Year2008      0.10097    0.23491    0.43  0.66748
## Year2009      0.20377    0.22737    0.90  0.37051
## Year2010     -0.03972    0.22740   -0.17  0.86141
## Year2011      0.06071    0.22896    0.27  0.79100
## Year2012      0.19207    0.23222    0.83  0.40852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.0429, Adjusted R-squared:  0.00624
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 57 weights are ~= 1. The remaining 540 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.195  0.871  0.951  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      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 2.149 1      1.466
## LastAuthorFemale  2.143 1      1.464
## Year              1.186 16      1.005

```


Residuals from first and last author



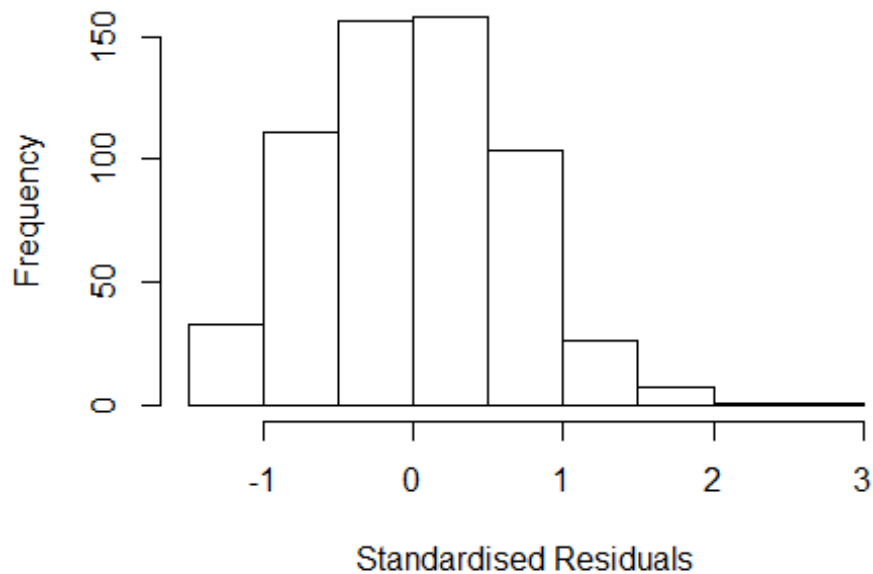
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 794 84858321141 3.398 2011      1200      3      2.515
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23844 -0.45364 -0.00331  0.45966  2.51536
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7881     0.2169   3.63 0.00031 ***
## FirstAuthorFemale1  0.0229     0.0995   0.23 0.81845
## LastAuthorFemale1  0.1054     0.1024   1.03 0.30398
## Year1997         -0.0245     0.2520  -0.10 0.92246
## Year1998          0.2496     0.2485   1.00 0.31554
## Year1999          0.4133     0.2805   1.47 0.14118
## Year2000         -0.0547     0.2795  -0.20 0.84482
## Year2001          0.0946     0.2478   0.38 0.70279
## Year2002          0.1734     0.2603   0.67 0.50570
## Year2003          0.0657     0.2668   0.25 0.80546
## Year2004          0.0922     0.2397   0.38 0.70078
## Year2005          0.0854     0.2462   0.35 0.72889
```

```

## Year2006          0.0102      0.2414      0.04  0.96634
## Year2007          0.3222      0.2342      1.38  0.16946
## Year2008          0.1138      0.2409      0.47  0.63684
## Year2009          0.2259      0.2328      0.97  0.33241
## Year2010         -0.0157      0.2319     -0.07  0.94588
## Year2011          0.0946      0.2330      0.41  0.68493
## Year2012          0.2211      0.2358      0.94  0.34898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.031, Adjusted R-squared:  0.000837
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.186  0.868  0.956  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      1.68e-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 1.09 1          1.044
## Year              1.09 16          1.003

```

Residuals from first author



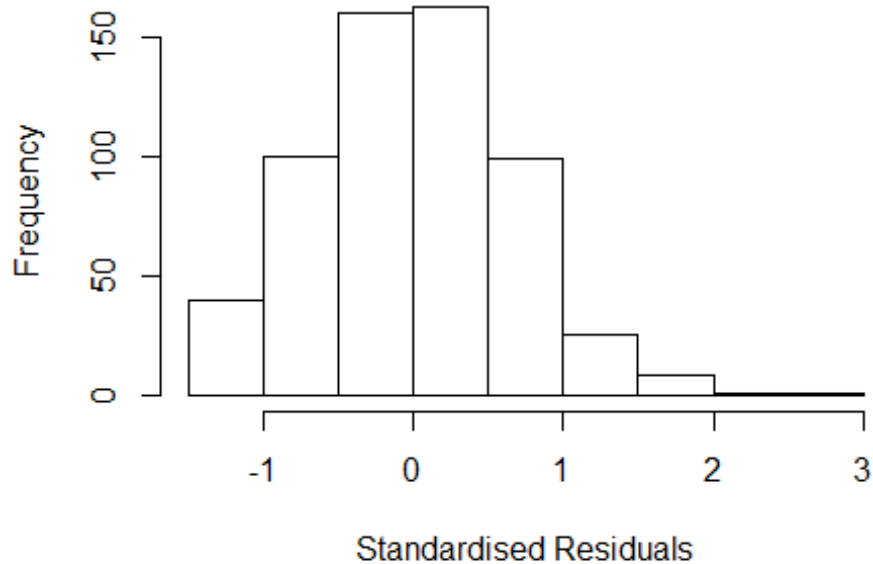
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 794 84858321141 3.398 2011    1200      3      2.515
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22162 -0.46102 -0.00503  0.45888  2.50863
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.78726    0.21361   3.69 0.00025 ***
## FirstAuthorFemale1 0.09168    0.07045   1.30 0.19367
## Year1997      -0.02382    0.24935  -0.10 0.92393
## Year1998       0.24760    0.24587   1.01 0.31434
## Year1999       0.43436    0.28027   1.55 0.12174
## Year2000      -0.05452    0.27280  -0.20 0.84166
## Year2001       0.09845    0.24419   0.40 0.68696
## Year2002       0.18390    0.25715   0.72 0.47480
## Year2003       0.05103    0.26340   0.19 0.84643
## Year2004       0.09676    0.23706   0.41 0.68330
## Year2005       0.08709    0.24224   0.36 0.71934
## Year2006       0.02601    0.23884   0.11 0.91331
```

```

## Year2007          0.32386    0.23146    1.40  0.16229
## Year2008          0.12609    0.23715    0.53  0.59515
## Year2009          0.23719    0.22993    1.03  0.30271
## Year2010         -0.00817    0.22850   -0.04  0.97148
## Year2011          0.10211    0.22909    0.45  0.65597
## Year2012          0.23268    0.23309    1.00  0.31857
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.714
## Multiple R-squared:  0.0288, Adjusted R-squared:  0.000315
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 544 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.192  0.868  0.954  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      1.68e-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 1.091 1          1.044
## Year            1.091 16          1.003

```

Residuals from last author



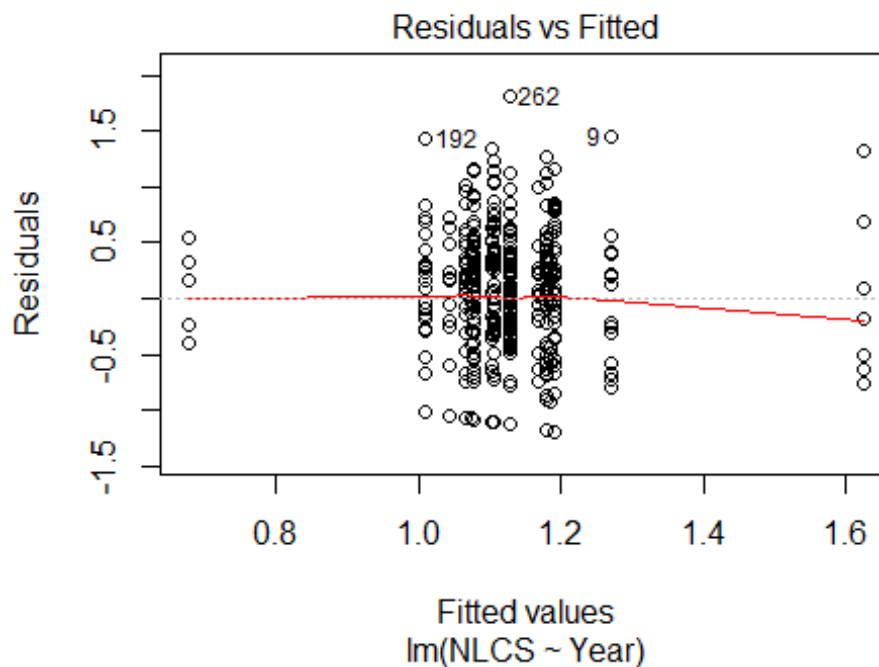
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 794 84858321141 3.398 2011      1200      3      2.515
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23318 -0.45698 -0.00239  0.45788  2.51202
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.79028    0.21799   3.63 0.00031 ***
## LastAuthorFemale1 0.12106    0.07302   1.66 0.09790 .
## Year1997      -0.02677    0.25312  -0.11 0.91581
## Year1998       0.24992    0.24931   1.00 0.31654
## Year1999       0.40899    0.28005   1.46 0.14472
## Year2000      -0.05454    0.28118  -0.19 0.84627
## Year2001       0.09472    0.24745   0.38 0.70201
## Year2002       0.17211    0.26162   0.66 0.51090
## Year2003       0.06943    0.26657   0.26 0.79462
## Year2004       0.09069    0.24075   0.38 0.70652
## Year2005       0.08482    0.24671   0.34 0.73111
## Year2006       0.00825    0.24238   0.03 0.97287
```

```

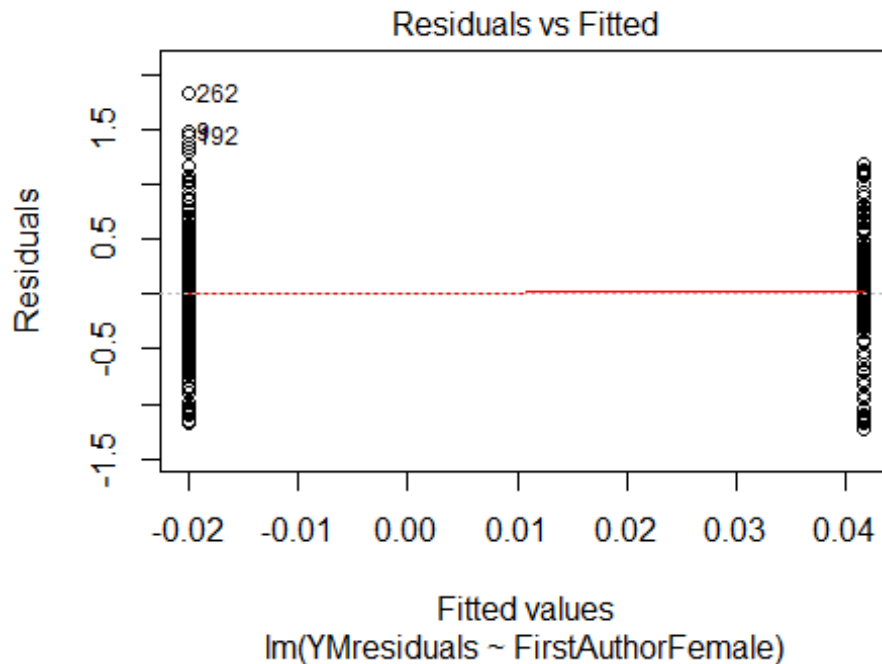
## Year2007      0.32184      0.23568      1.37  0.17260
## Year2008      0.11312      0.24104      0.47  0.63903
## Year2009      0.22422      0.23370      0.96  0.33774
## Year2010     -0.01601      0.23226     -0.07  0.94506
## Year2011      0.09569      0.23286      0.41  0.68127
## Year2012      0.22024      0.23678      0.93  0.35268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0309, Adjusted R-squared:  0.00248
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.187  0.868  0.956  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      1.68e-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: 597"
## [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
##   19   13   15   14   19    8   18   23   19   39   43   37   59   53   69
## 2011 2012
##   55   74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15    7   13   12   16    6   14   19   15   29   40   30   41   38   53
## 2011 2012

```

```
## 43 54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 7 11 11 14 5 11 16 13 28 33 29 39 35 50
## 2011 2012
## 41 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 = 12, df = 16, p-value = 0.7
```



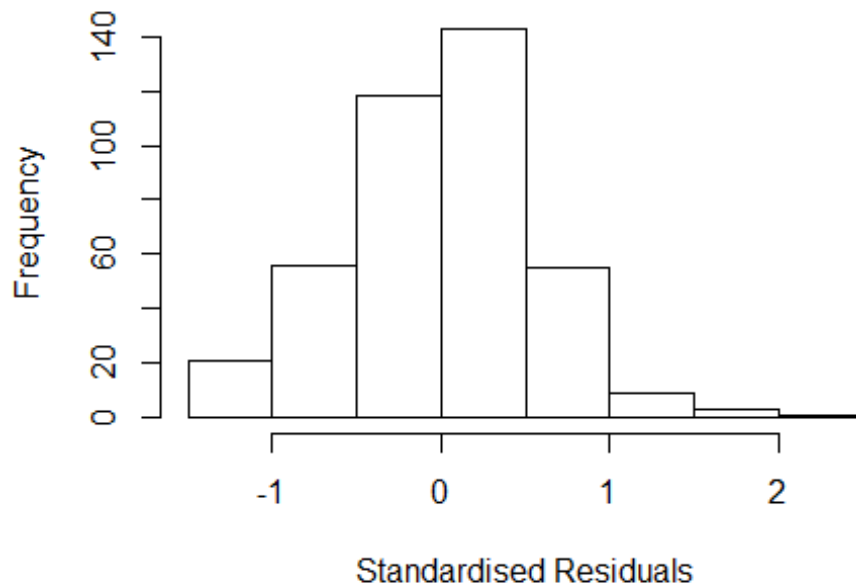
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 2.48716219977362"
## [1] "Male first author team size 2018 geometric mean: 2.12493235518678"
## 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 = 590, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.3038414060869"
## [1] "Male last author team size 2018 geometric mean: 2.21044047515958"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	2.804	1	1.674
LastAuthorFemale	1.711	1	1.308
UniqueAuthors	5.636	4	1.241
Year	6.127	16	1.058

Residuals from first and last author and team size



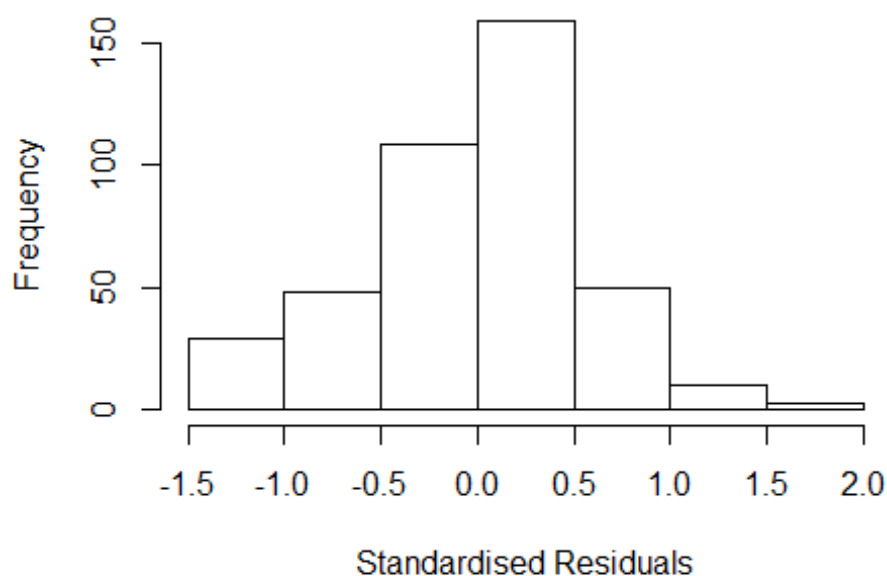
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4147 -0.3604 0.0191 0.3526 2.0298
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01947 0.16179 6.30 8.1e-10 ***
## FirstAuthorFemale1 -0.03590 0.07167 -0.50 0.61668
## LastAuthorFemale1 0.21903 0.07820 2.80 0.00536 **
## UniqueAuthors2 0.18935 0.07297 2.59 0.00983 **
## UniqueAuthors3 0.32361 0.07323 4.42 1.3e-05 ***
## UniqueAuthors4 0.35021 0.16046 2.18 0.02968 *
## UniqueAuthors5 -0.37181 0.09832 -3.78 0.00018 ***
## Year1997 0.20722 0.28840 0.72 0.47289
## Year1998 -0.10840 0.19975 -0.54 0.58767
## Year1999 -0.01313 0.22349 -0.06 0.95318
```

```

## Year2000          0.00145    0.18848    0.01  0.99387
## Year2001         -0.63022    0.28456   -2.21  0.02737 *
## Year2002         -0.09466    0.26067   -0.36  0.71669
## Year2003         -0.10870    0.21323   -0.51  0.61052
## Year2004         -0.07072    0.19640   -0.36  0.71899
## Year2005         -0.22550    0.19563   -1.15  0.24977
## Year2006         -0.11422    0.17542   -0.65  0.51535
## Year2007         -0.00431    0.19216   -0.02  0.98213
## Year2008          0.03546    0.18744    0.19  0.85004
## Year2009         -0.18917    0.18080   -1.05  0.29607
## Year2010         -0.10837    0.17242   -0.63  0.53003
## Year2011         -0.14072    0.18266   -0.77  0.44155
## Year2012         -0.07615    0.18140   -0.42  0.67486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0551
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 367 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0971 0.8650 0.9450 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      2.46e-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 1.406 1      1.186
## LastAuthorFemale  1.399 1      1.183
## Year              1.551 16      1.014

```

Residuals from first and last author



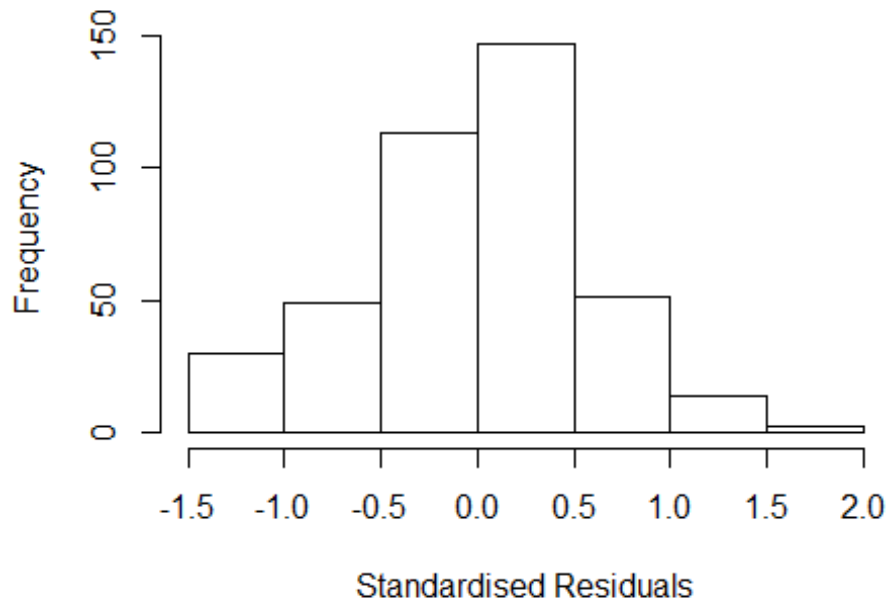
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3815 -0.3778 0.0484 0.3311 1.8536
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16076 0.17243 6.73 6.1e-11 ***
## FirstAuthorFemale1 -0.02979 0.07216 -0.41 0.6800
## LastAuthorFemale1 0.22014 0.08031 2.74 0.0064 **
## Year1997 0.27023 0.31503 0.86 0.3915
## Year1998 -0.12757 0.20267 -0.63 0.5294
## Year1999 -0.10063 0.22713 -0.44 0.6580
## Year2000 -0.02640 0.19624 -0.13 0.8931
## Year2001 -0.60438 0.26319 -2.30 0.0222 *
## Year2002 -0.13860 0.26793 -0.52 0.6052
## Year2003 -0.06401 0.22622 -0.28 0.7774
## Year2004 -0.04631 0.21089 -0.22 0.8263
## Year2005 -0.19842 0.20492 -0.97 0.3335
```

```

## Year2006      -0.07935    0.18602   -0.43    0.6700
## Year2007      0.03043    0.20465    0.15    0.8819
## Year2008      0.00772    0.19470    0.04    0.9684
## Year2009     -0.18337    0.19489   -0.94    0.3473
## Year2010     -0.10474    0.18758   -0.56    0.5769
## Year2011     -0.12523    0.19761   -0.63    0.5266
## Year2012     -0.08117    0.19791   -0.41    0.6820
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0537, Adjusted R-squared:  0.00972
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 368 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.864  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      2.46e-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 1.187 1      1.089
## Year      1.187 16      1.005

```

Residuals from first author



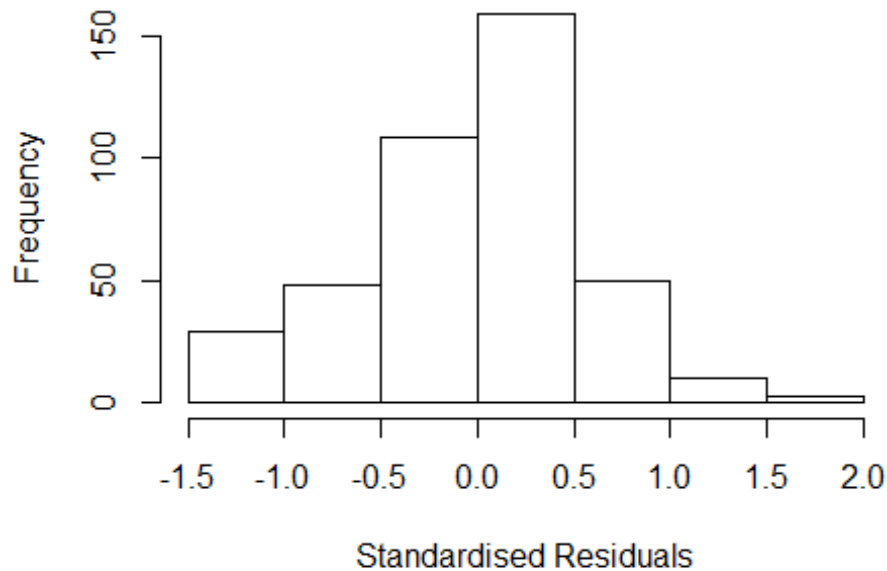
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2625 -0.3583 0.0474 0.3409 1.8282
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17463 0.16154 7.27 2e-12 ***
## FirstAuthorFemale1 0.06420 0.06773 0.95 0.344
## Year1997 0.34483 0.33863 1.02 0.309
## Year1998 -0.07333 0.19436 -0.38 0.706
## Year1999 -0.09764 0.22203 -0.44 0.660
## Year2000 0.01096 0.18857 0.06 0.954
## Year2001 -0.57430 0.23584 -2.44 0.015 *
## Year2002 -0.07470 0.27038 -0.28 0.782
## Year2003 -0.08520 0.21855 -0.39 0.697
## Year2004 -0.02535 0.20293 -0.12 0.901
## Year2005 -0.19403 0.19694 -0.99 0.325
## Year2006 -0.06782 0.17746 -0.38 0.703
```

```

## Year2007          0.02367      0.19321      0.12      0.903
## Year2008          0.00148      0.18694      0.01      0.994
## Year2009         -0.15625      0.18418     -0.85      0.397
## Year2010         -0.11752      0.17794     -0.66      0.509
## Year2011         -0.07835      0.18333     -0.43      0.669
## Year2012         -0.07660      0.18633     -0.41      0.681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0316, Adjusted R-squared:  -0.0108
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 381 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.246  0.867  0.958  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      2.46e-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 1.196 1      1.094
## Year            1.196 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
## -1.3911 -0.3769 0.0514 0.3301 1.8618
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1514 0.1668 6.90 2.1e-11 ***
## LastAuthorFemale1 0.2056 0.0751 2.74 0.0065 **
## Year1997 0.2852 0.3116 0.92 0.3606
## Year1998 -0.1228 0.2003 -0.61 0.5400
## Year1999 -0.0946 0.2238 -0.42 0.6727
## Year2000 -0.0191 0.1930 -0.10 0.9212
## Year2001 -0.5923 0.2573 -2.30 0.0219 *
## Year2002 -0.1285 0.2643 -0.49 0.6272
## Year2003 -0.0652 0.2246 -0.29 0.7716
## Year2004 -0.0390 0.2075 -0.19 0.8508
## Year2005 -0.1928 0.2024 -0.95 0.3415
## Year2006 -0.0782 0.1847 -0.42 0.6723
```

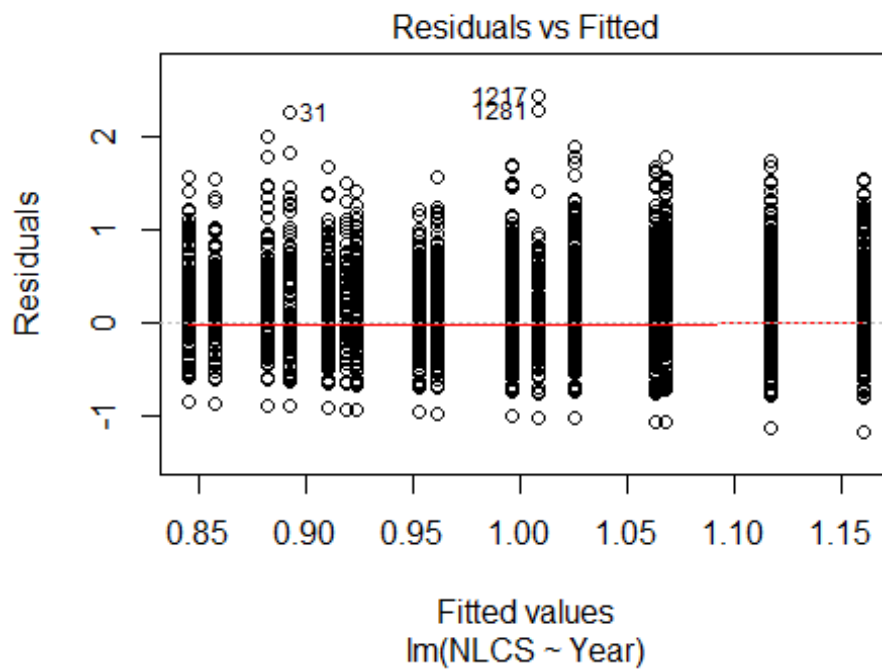
```

## Year2007          0.0341      0.2023      0.17      0.8664
## Year2008          0.0102      0.1927      0.05      0.9578
## Year2009         -0.1814      0.1935     -0.94      0.3492
## Year2010         -0.1050      0.1864     -0.56      0.5736
## Year2011         -0.1193      0.1928     -0.62      0.5365
## Year2012         -0.0757      0.1944     -0.39      0.6971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.53
## Multiple R-squared:  0.0534, Adjusted R-squared:  0.012
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.191  0.864  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      2.46e-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: 406"
## [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
##  202  249  303  289  288  195  276  283  333  285  332  362  479  466  567
## 2011 2012
##  509  595
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  153  192  223  207  193  101  209  205  272  213  264  282  383  371  462
## 2011 2012

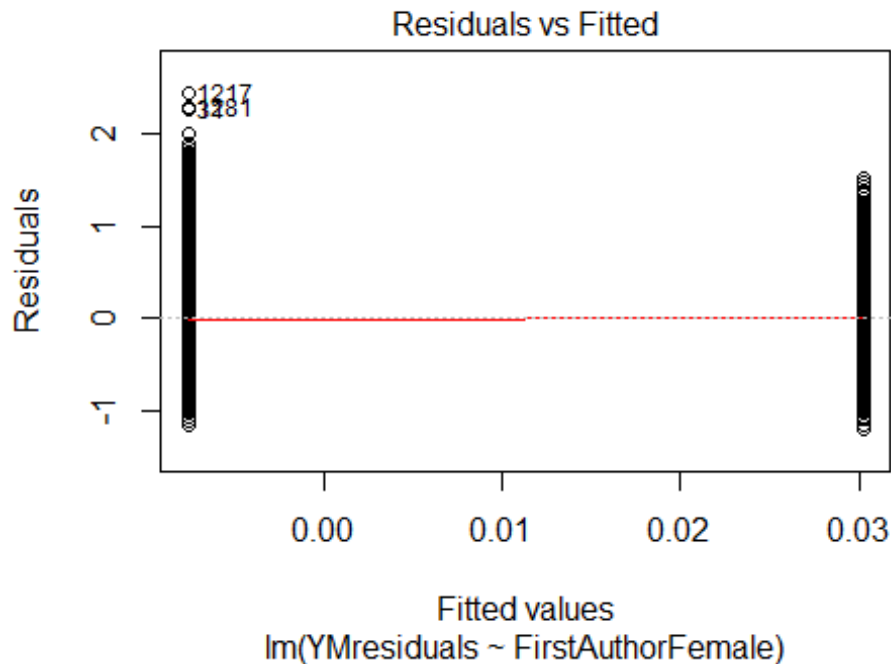
```



```
## 397 467
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 143 187 211 199 183 94 194 183 251 191 239 247 342 336 409
## 2011 2012
## 356 416
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 86, df = 16, p-value = 2e-11
```

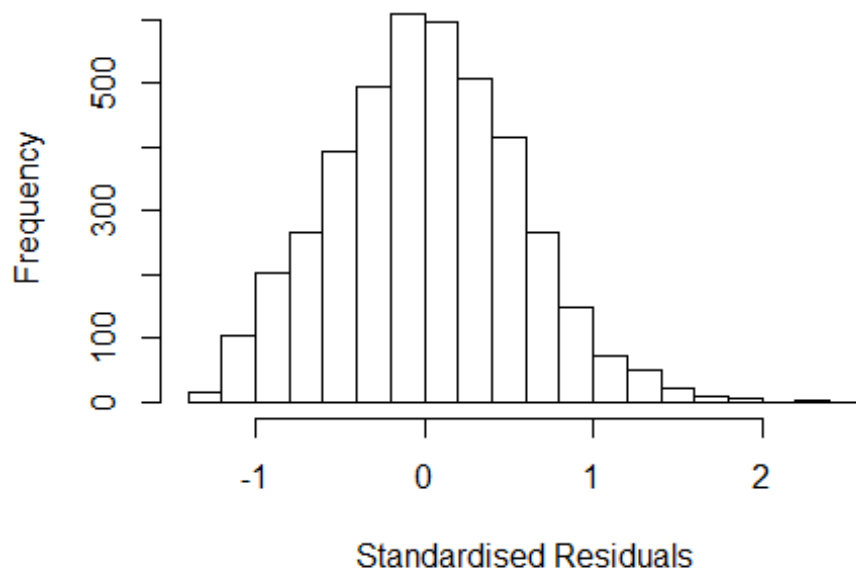


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 1.86852546755803"
## [1] "Male first author team size 2018 geometric mean: 1.66374684201627"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.75884741792083"
## [1] "Male last author team size 2018 geometric mean: 1.70840218694706"
##
## 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 1.464 1      1.210
## LastAuthorFemale  1.452 1      1.205
## UniqueAuthors    1.106 4      1.013
## Year             1.119 16      1.004
```

Residuals from first and last author and team size



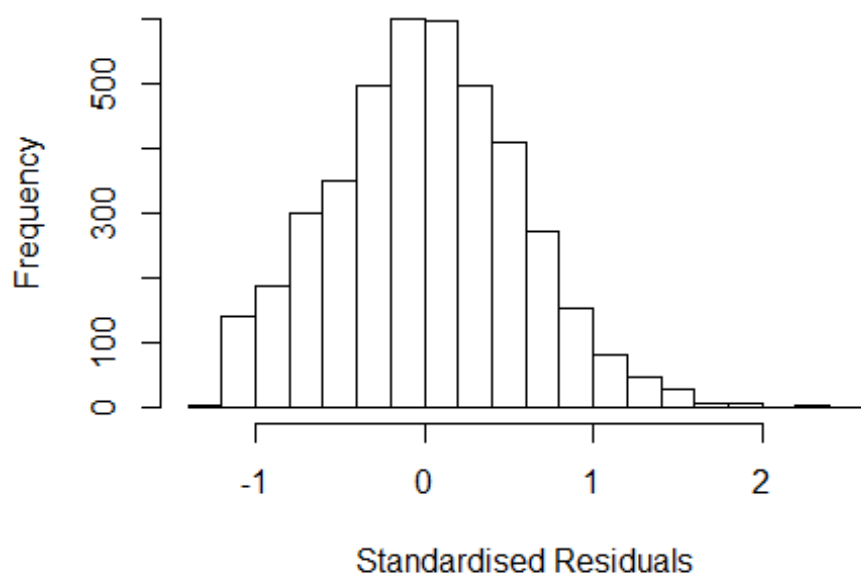
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1217 0034565913 3.441 2000      2002      2      2.529
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.32596 -0.36694  0.00363  0.37697  2.52902
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77782    0.04185   18.59 < 2e-16 ***
## FirstAuthorFemale1 -0.00691    0.02586   -0.27  0.78921
## LastAuthorFemale1  0.06524    0.02635    2.48  0.01332 *
## UniqueAuthors2    0.13596    0.02017    6.74  1.8e-11 ***
## UniqueAuthors3    0.20982    0.02912    7.21  6.8e-13 ***
## UniqueAuthors4    0.25552    0.06015    4.25  2.2e-05 ***
## UniqueAuthors5    0.47907    0.11731    4.08  4.5e-05 ***
## Year1997          0.02470    0.05768    0.43  0.66847
## Year1998         -0.00568    0.05580   -0.10  0.91892
## Year1999          0.00203    0.05438    0.04  0.97019
```

```

## Year2000          0.13416      0.05504      2.44  0.01484 *
## Year2001          0.08208      0.07272      1.13  0.25905
## Year2002          0.05492      0.05549      0.99  0.32234
## Year2003          0.02927      0.05746      0.51  0.61053
## Year2004          0.09157      0.05210      1.76  0.07891 .
## Year2005          0.12194      0.05629      2.17  0.03033 *
## Year2006          0.08951      0.05252      1.70  0.08843 .
## Year2007          0.13615      0.05366      2.54  0.01121 *
## Year2008          0.12392      0.05350      2.32  0.02060 *
## Year2009          0.19543      0.05332      3.67  0.00025 ***
## Year2010          0.23145      0.05317      4.35  1.4e-05 ***
## Year2011          0.27307      0.05403      5.05  4.5e-07 ***
## Year2012          0.17839      0.05305      3.36  0.00078 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0561, Adjusted R-squared:  0.0511
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 361 weights are ~= 1. The remaining 3820 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0031 0.8750 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          2.39e-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 1.455 1          1.206
## LastAuthorFemale 1.443 1          1.201
## Year              1.045 16          1.001

```

Residuals from first and last author



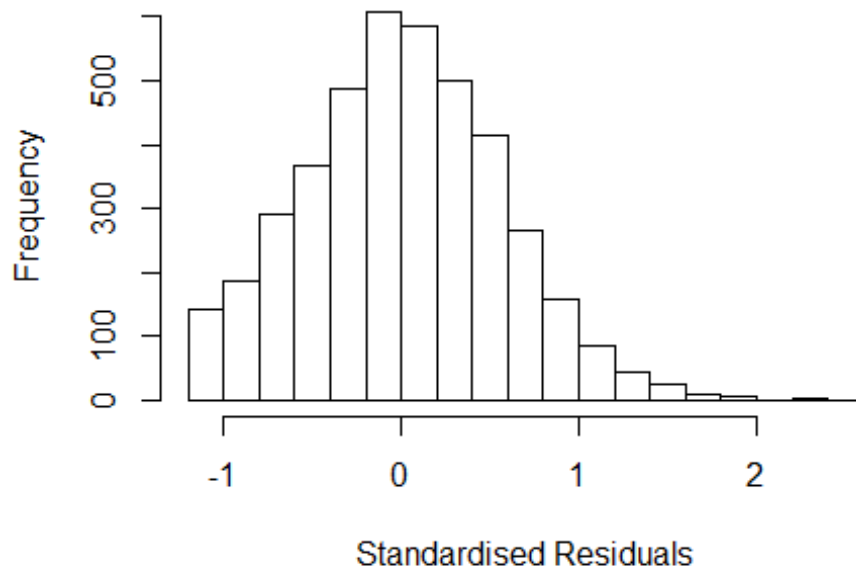
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22183 -0.38183  0.00334  0.38297  2.46630
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82803    0.04268   19.40  < 2e-16 ***
## FirstAuthorFemale1 0.00677    0.02621    0.26  0.7961
## LastAuthorFemale1 0.08003    0.02676    2.99  0.0028 **
## Year1997         0.01529    0.05910    0.26  0.7959
## Year1998        -0.00702    0.05657   -0.12  0.9013
## Year1999         0.00658    0.05574    0.12  0.9060
## Year2000         0.14668    0.05587    2.63  0.0087 **
## Year2001         0.08539    0.07419    1.15  0.2498
## Year2002         0.05939    0.05699    1.04  0.2974
## Year2003         0.04554    0.05806    0.78  0.4329
## Year2004         0.10480    0.05312    1.97  0.0486 *
## Year2005         0.14566    0.05697    2.56  0.0106 *
```

```

## Year2006          0.10164      0.05354      1.90      0.0577 .
## Year2007          0.15983      0.05429      2.94      0.0033 **
## Year2008          0.14346      0.05460      2.63      0.0086 **
## Year2009          0.21636      0.05477      3.95      7.9e-05 ***
## Year2010          0.25447      0.05404      4.71      2.6e-06 ***
## Year2011          0.30700      0.05521      5.56      2.9e-08 ***
## Year2012          0.20215      0.05427      3.72      0.0002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.032, Adjusted R-squared:  0.0278
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 339 weights are ~= 1. The remaining 3842 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0158 0.8730 0.9510 0.9090 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      2.39e-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 1.033 1      1.016
## Year      1.033 16      1.001

```

Residuals from first author



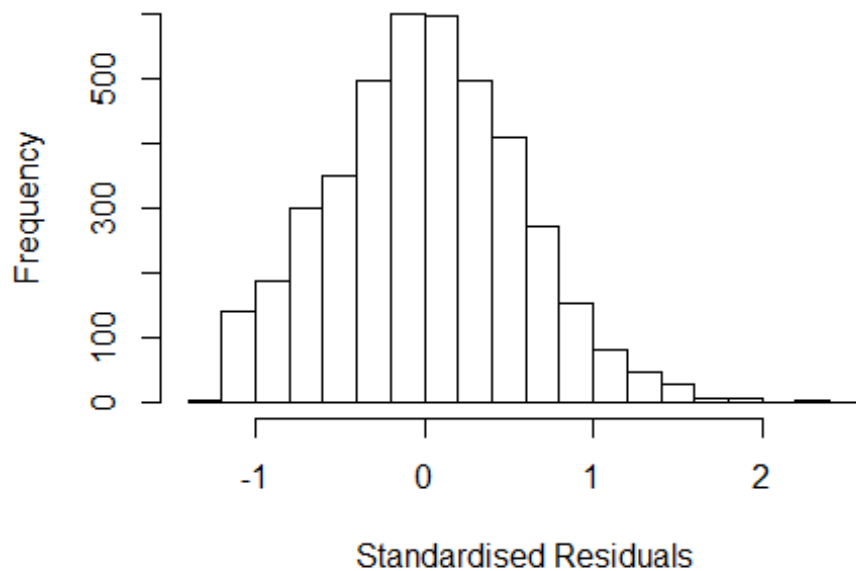
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19188 -0.38393 0.00072 0.38224 2.46200
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8299 0.0428 19.40 < 2e-16 ***
## FirstAuthorFemale1 0.0496 0.0220 2.26 0.02410 *
## Year1997 0.0191 0.0595 0.32 0.74772
## Year1998 -0.0048 0.0567 -0.08 0.93252
## Year1999 0.0094 0.0558 0.17 0.86616
## Year2000 0.1491 0.0560 2.66 0.00778 **
## Year2001 0.0892 0.0740 1.21 0.22774
## Year2002 0.0644 0.0571 1.13 0.25931
## Year2003 0.0522 0.0581 0.90 0.36940
## Year2004 0.1114 0.0531 2.10 0.03586 *
## Year2005 0.1505 0.0572 2.63 0.00853 **
## Year2006 0.1040 0.0538 1.93 0.05325 .
```

```

## Year2007          0.1660      0.0542      3.06  0.00223 **
## Year2008          0.1507      0.0545      2.76  0.00573 **
## Year2009          0.2221      0.0548      4.05  5.2e-05 ***
## Year2010          0.2618      0.0541      4.84  1.3e-06 ***
## Year2011          0.3123      0.0554      5.64  1.8e-08 ***
## Year2012          0.2088      0.0544      3.84  0.00013 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0297, Adjusted R-squared:  0.0257
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 348 weights are ~= 1. The remaining 3833 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0175 0.8750 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      2.39e-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 1.025 1      1.012
## Year            1.025 16      1.001

```


Residuals from last author



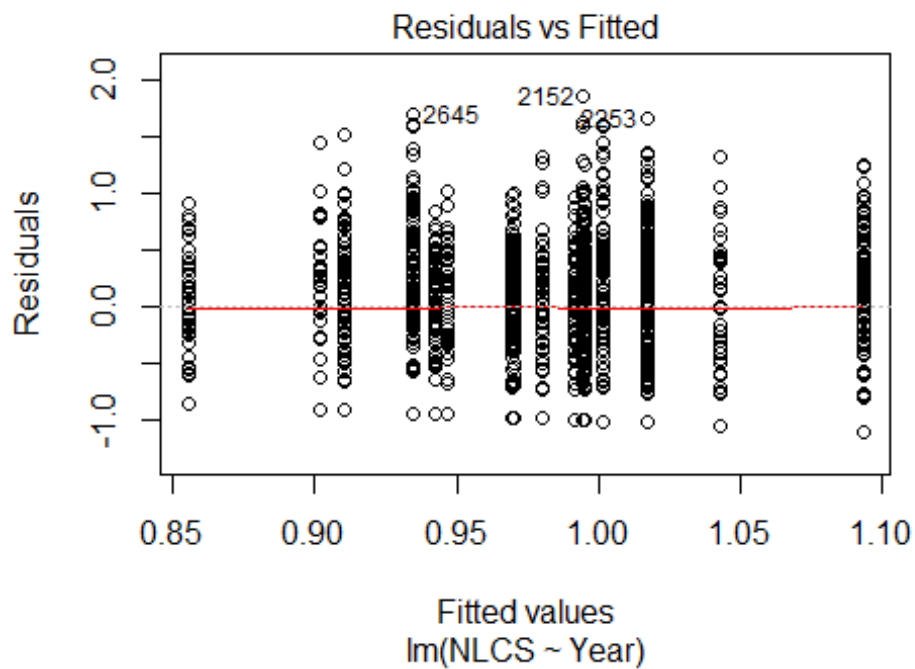
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21970 -0.38112 0.00334 0.38296 2.46562
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82842 0.04269 19.41 < 2e-16 ***
## LastAuthorFemale1 0.08376 0.02255 3.71 0.00021 ***
## Year1997 0.01507 0.05910 0.26 0.79871
## Year1998 -0.00705 0.05657 -0.12 0.90079
## Year1999 0.00677 0.05571 0.12 0.90334
## Year2000 0.14696 0.05582 2.63 0.00851 **
## Year2001 0.08563 0.07419 1.15 0.24849
## Year2002 0.05945 0.05697 1.04 0.29678
## Year2003 0.04583 0.05802 0.79 0.42961
## Year2004 0.10494 0.05311 1.98 0.04825 *
## Year2005 0.14589 0.05693 2.56 0.01042 *
## Year2006 0.10219 0.05341 1.91 0.05580 .
```

```

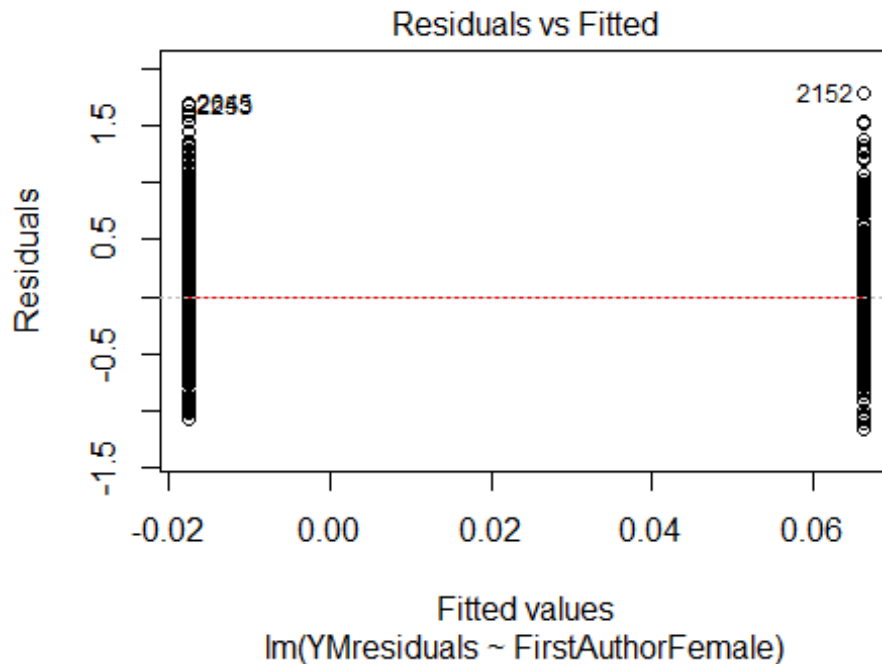
## Year2007      0.16024      0.05423      2.95  0.00315 **
## Year2008      0.14361      0.05458      2.63  0.00854 **
## Year2009      0.21664      0.05472      3.96  7.7e-05 ***
## Year2010      0.25468      0.05401      4.72  2.5e-06 ***
## Year2011      0.30752      0.05507      5.58  2.5e-08 ***
## Year2012      0.20262      0.05419      3.74  0.00019 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.032, Adjusted R-squared:  0.028
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 341 weights are ~= 1. The remaining 3840 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0158 0.8730 0.9510 0.9090 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      2.39e-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: 4181"
## [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
##   54   49   79   80   80   88   82   78  127  129  133  162  200  179  241
## 2011 2012
##  215  269
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   40   38   62   57   63   59   66   64  102   98  104  120  137  134  189
## 2011 2012

```

```
## 154 191
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 32 37 59 55 58 54 59 52 88 87 89 103 120 110 153
## 2011 2012
## 130 167
## [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 = 4e-06
```

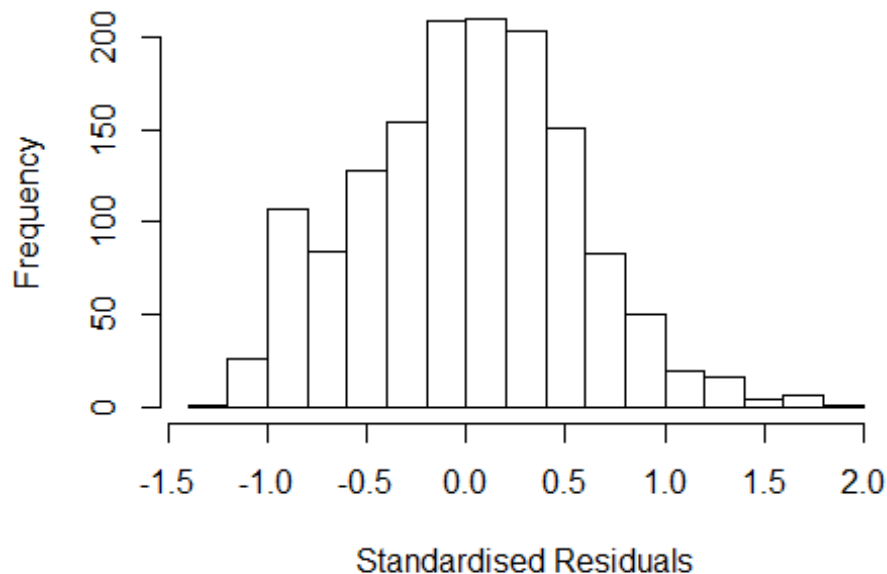


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.4, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 2.05390600781467"
## [1] "Male first author team size 2018 geometric mean: 1.79939672584591"
##
## 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: 1.99095685408765"
## [1] "Male last author team size 2018 geometric mean: 1.82689213196105"
##
## 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 1.250 1          1.118
## LastAuthorFemale  1.225 1          1.107
## UniqueAuthors    1.277 4          1.031
## Year              1.310 16         1.008
```

Residuals from first and last author and team size



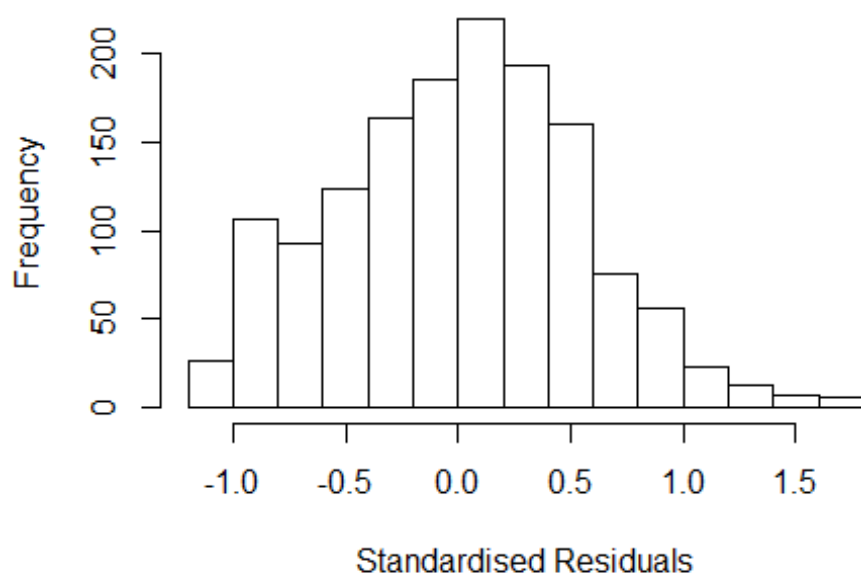
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2660 -0.3876 0.0201 0.3791 1.8218
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.88199 0.10866 8.12 1e-15 ***
## FirstAuthorFemale1 -0.01050 0.04293 -0.24 0.80677
## LastAuthorFemale1 0.13254 0.04731 2.80 0.00516 **
## UniqueAuthors2 0.10873 0.03389 3.21 0.00136 **
## UniqueAuthors3 0.16451 0.04556 3.61 0.00032 ***
## UniqueAuthors4 0.04391 0.10135 0.43 0.66493
## UniqueAuthors5 -0.09576 0.20647 -0.46 0.64285
## Year1997 -0.05092 0.14529 -0.35 0.72604
## Year1998 -0.00639 0.12376 -0.05 0.95881
## Year1999 -0.09398 0.12452 -0.75 0.45051
```

```

## Year2000          0.03247    0.12394    0.26  0.79338
## Year2001          0.01526    0.13637    0.11  0.91093
## Year2002          0.04143    0.12735    0.33  0.74497
## Year2003         -0.02084    0.12621   -0.17  0.86886
## Year2004         -0.00787    0.11946   -0.07  0.94750
## Year2005         -0.06366    0.12528   -0.51  0.61142
## Year2006          0.00381    0.12311    0.03  0.97534
## Year2007         -0.00437    0.11802   -0.04  0.97048
## Year2008         -0.01174    0.12271   -0.10  0.92377
## Year2009          0.09743    0.12175    0.80  0.42369
## Year2010         -0.01523    0.12064   -0.13  0.89957
## Year2011          0.02490    0.12435    0.20  0.84130
## Year2012         -0.07976    0.12025   -0.66  0.50728
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0166
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1335 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.246  0.876  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      6.88e-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 1.213 1      1.101
## LastAuthorFemale  1.219 1      1.104
## Year              1.065 16      1.002

```

Residuals from first and last author



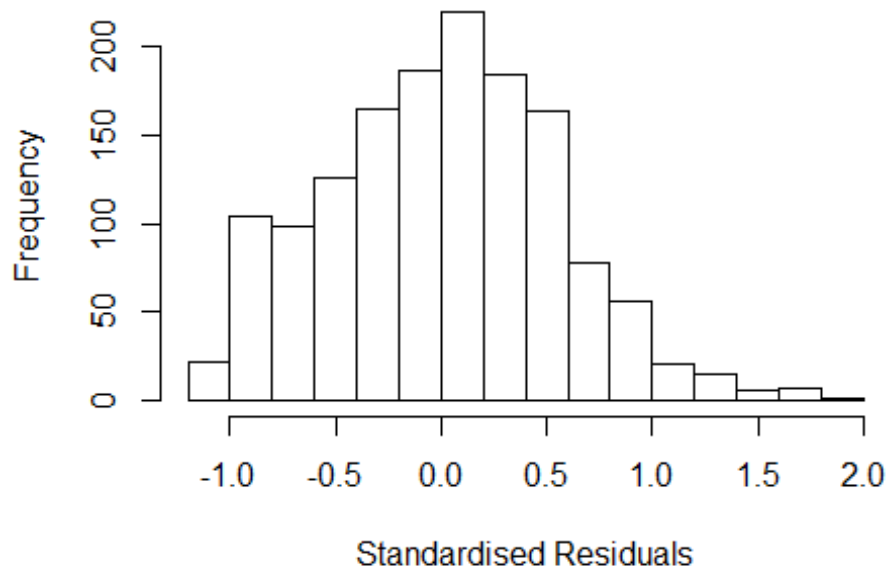
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.184 -0.375 0.022 0.369 1.771
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.913358 0.110913 8.23 4e-16 ***
## FirstAuthorFemale1 0.003445 0.042838 0.08 0.9359
## LastAuthorFemale1 0.133233 0.048201 2.76 0.0058 **
## Year1997 -0.033171 0.149649 -0.22 0.8246
## Year1998 -0.001716 0.125390 -0.01 0.9891
## Year1999 -0.084119 0.126481 -0.67 0.5061
## Year2000 0.058023 0.125389 0.46 0.6436
## Year2001 0.035147 0.134790 0.26 0.7943
## Year2002 0.064761 0.129952 0.50 0.6183
## Year2003 -0.000335 0.128390 0.00 0.9979
## Year2004 0.015219 0.121030 0.13 0.9000
## Year2005 -0.020279 0.126380 -0.16 0.8725
```

```

## Year2006      0.049063  0.124805  0.39  0.6943
## Year2007      0.034807  0.119173  0.29  0.7703
## Year2008      0.027474  0.123694  0.22  0.8243
## Year2009      0.134368  0.123310  1.09  0.2760
## Year2010      0.027756  0.121533  0.23  0.8194
## Year2011      0.052173  0.126152  0.41  0.6792
## Year2012     -0.052092  0.121899 -0.43  0.6692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.00583
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1337 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.277  0.876  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      6.88e-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 1.031 1      1.015
## Year              1.031 16      1.001

```


Residuals from first author



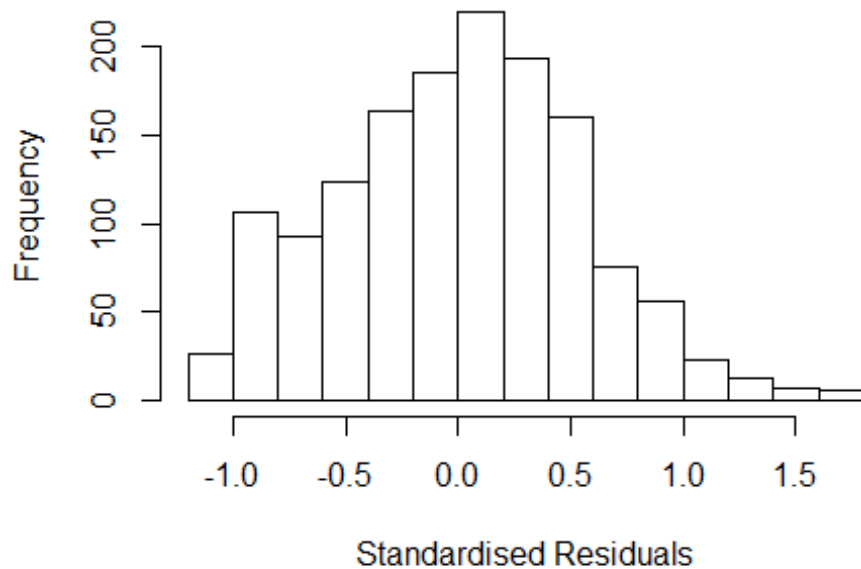
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1277 -0.3775 0.0165 0.3809 1.8318
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92429 0.11163 8.28 2.8e-16 ***
## FirstAuthorFemale1 0.06167 0.03966 1.55 0.12
## Year1997 -0.03523 0.15159 -0.23 0.82
## Year1998 0.00185 0.12676 0.01 0.99
## Year1999 -0.08470 0.12739 -0.66 0.51
## Year2000 0.05255 0.12649 0.42 0.68
## Year2001 0.03485 0.13479 0.26 0.80
## Year2002 0.06539 0.13156 0.50 0.62
## Year2003 -0.00588 0.12963 -0.05 0.96
## Year2004 0.01946 0.12234 0.16 0.87
## Year2005 -0.02202 0.12735 -0.17 0.86
## Year2006 0.04325 0.12588 0.34 0.73
```

```

## Year2007          0.03735    0.12024    0.31    0.76
## Year2008          0.03448    0.12494    0.28    0.78
## Year2009          0.14170    0.12424    1.14    0.25
## Year2010          0.03125    0.12262    0.25    0.80
## Year2011          0.05340    0.12670    0.42    0.67
## Year2012         -0.05324    0.12284   -0.43    0.66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0109, Adjusted R-squared:  -0.000783
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1330 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.880  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      6.88e-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 1.036 1      1.018
## Year              1.036 16      1.001

```

Residuals from last author



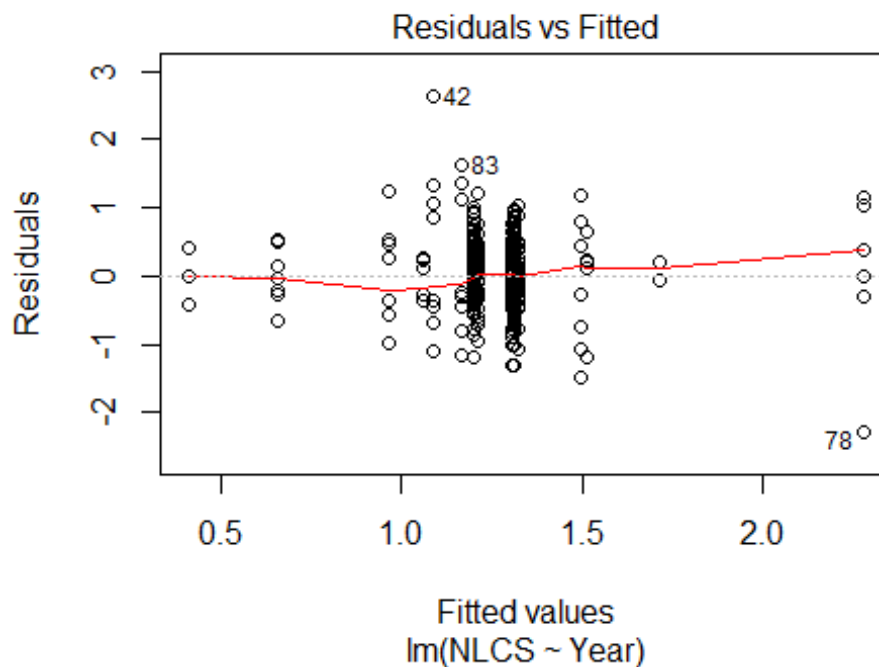
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1831 -0.3725 0.0215 0.3704 1.7725
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.913836 0.111001 8.23 4.1e-16 ***
## LastAuthorFemale1 0.134881 0.044461 3.03 0.0025 **
## Year1997 -0.033574 0.149751 -0.22 0.8226
## Year1998 -0.002033 0.125509 -0.02 0.9871
## Year1999 -0.084170 0.126480 -0.67 0.5059
## Year2000 0.057951 0.125389 0.46 0.6440
## Year2001 0.034784 0.134950 0.26 0.7966
## Year2002 0.064576 0.129983 0.50 0.6194
## Year2003 -0.000335 0.128330 0.00 0.9979
## Year2004 0.015107 0.121060 0.12 0.9007
## Year2005 -0.020381 0.126421 -0.16 0.8719
## Year2006 0.049259 0.124565 0.40 0.6926
```

```

## Year2007      0.034815   0.119138   0.29   0.7702
## Year2008      0.027323   0.123758   0.22   0.8253
## Year2009      0.134343   0.123293   1.09   0.2761
## Year2010      0.027757   0.121512   0.23   0.8193
## Year2011      0.051992   0.126155   0.41   0.6803
## Year2012     -0.052025   0.121850  -0.43   0.6695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0181, Adjusted R-squared:  0.00651
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1334 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.276  0.876  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      6.88e-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: 1453"
## [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
##   13   17   19   12    9    8    8    9   20   21   24   83   71   74   82
## 2011 2012
##  115  124
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    7   11    8    6    3    6    4    5    8    5   45   35   50   53
## 2011 2012

```

```
## 61 64
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 6 11 7 6 3 6 4 4 6 4 43 29 41 47
## 2011 2012
## 51 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 = 87, df = 16, p-value = 1e-11
```



```
##
## 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: 3.5928324311208"
## [1] "Male first author team size 2018 geometric mean: 2.85410327813164"

## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.90993667492986"
## [1] "Male last author team size 2018 geometric mean: 3.05479989662265"

## 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 = 0.6
## 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 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"

## 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 outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2001'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.

## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate

## [1] "Sample size for the above analysis: 324"
## [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]"
##
## 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    2   13    2   59    3    5    7    7    6    6   10    7    3    3    9
##
## 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    2    3    1    0    3    3    5    6    4    4    6    6    2    2    8
##
## 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    2    1    0    2    3    5    6    4    4    6    6    2    2    7
## [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: 2.59326058903878"

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

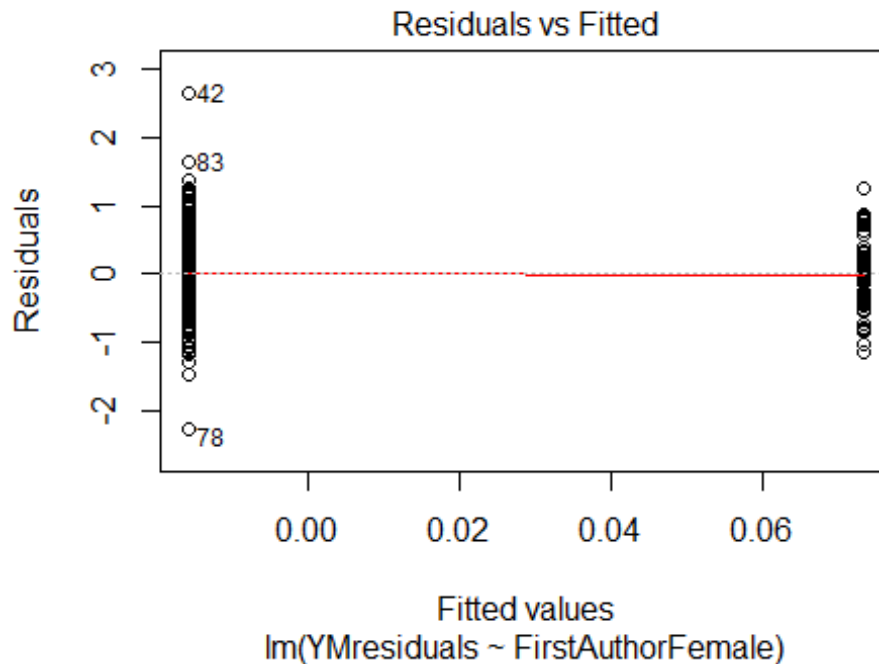
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.44224957030741"
## [1] "Male last author team size 2018 geometric mean: 2.56435912012936"

## 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 = 0.2
## 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 outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year1999'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.

```



```
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##                               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -1.651e+14  1             NaN
## LastAuthorFemale  -5.697e+14  1             NaN
## Year               -1.372e+28 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 + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.03e+00 -2.88e-01  1.14e-16  2.36e-01  1.61e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.14e-16   2.23e-08    0.00   1.0000
## FirstAuthorFemale1  2.55e-01   2.79e-01    0.91   0.3670
## LastAuthorFemale1  6.38e-02   2.43e-01    0.26   0.7942
## Year1999         2.03e+00   2.89e-01    7.03  3.5e-08 ***
## Year2000         3.05e-01   2.79e-01    1.09   0.2820
```



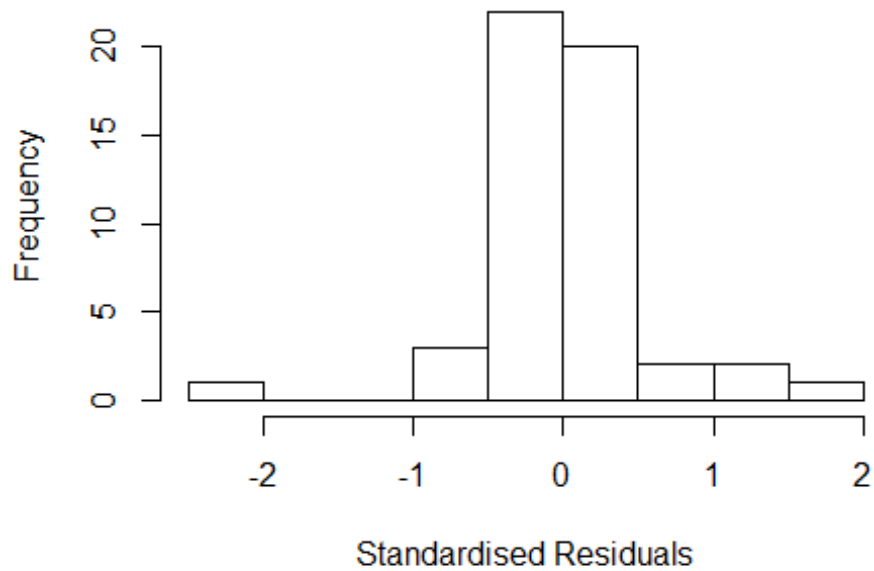
```

## Year2002          5.93e-01    8.67e-02    6.84  6.1e-08 ***
## Year2003          6.50e-01    3.31e-01    1.96  0.0579 .
## Year2004          3.28e-01    2.15e-01    1.53  0.1359
## Year2005          3.81e-01    1.51e-01    2.53  0.0161 *
## Year2006          3.38e-01    4.00e-01    0.84  0.4043
## Year2007          8.08e-01    9.58e-02    8.43  6.0e-10 ***
## Year2008          3.04e-01    1.96e-01    1.55  0.1307
## Year2009          2.72e-01    2.26e-01    1.20  0.2379
## Year2010          2.44e-01    1.81e-01    1.35  0.1873
## Year2011          1.29e+00    4.11e-01    3.14  0.0035 **
## Year2012          7.28e-01    2.33e-01    3.13  0.0035 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.423, Adjusted R-squared:  0.176
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 44 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0245 0.9310 0.9610 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      1.96e-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 outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year1999'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.

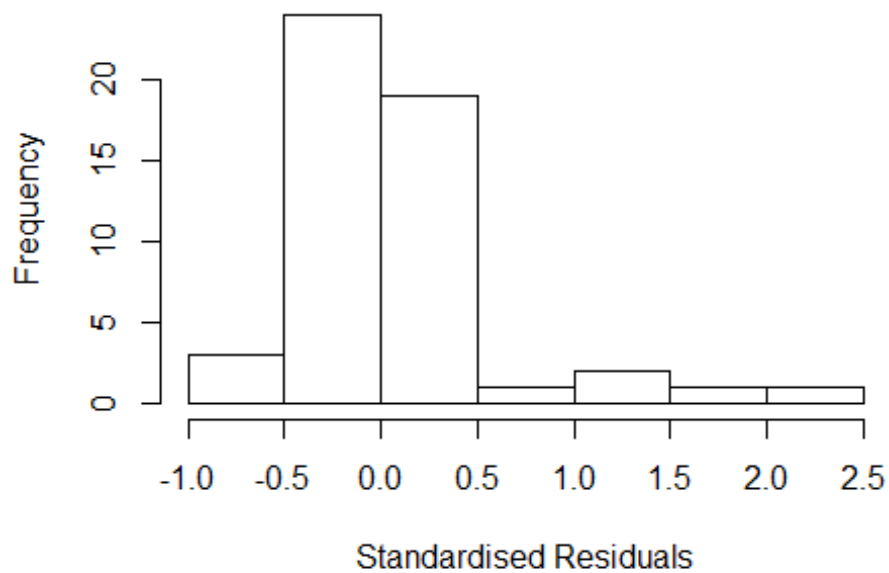
```

Residuals from first and last author



##		GVIF	Df	$GVIF^{(1/(2*Df))}$
##	FirstAuthorFemale	-1.519e+14	1	NaN
##	Year	-1.519e+14	13	NaN

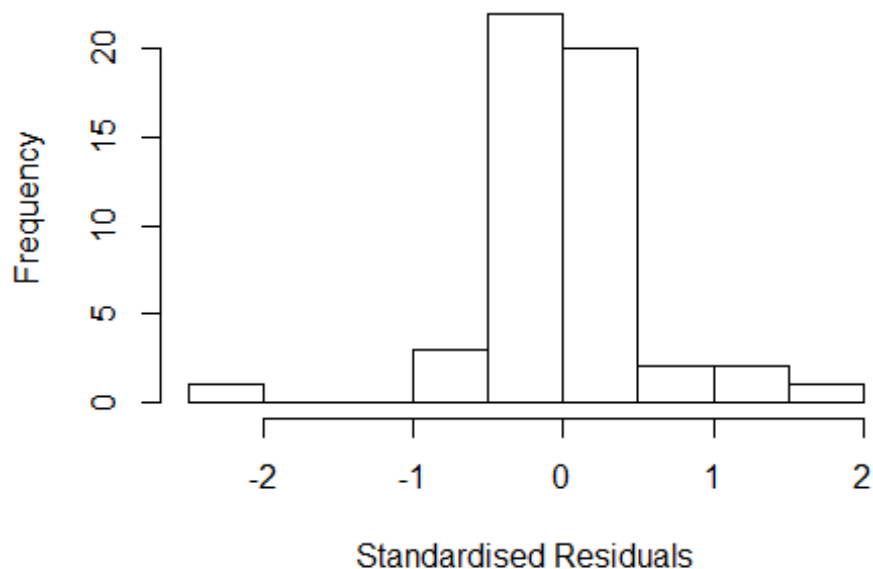
Residuals from first author



```
## [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
## -7.54e-01 -2.52e-01 -1.06e-16 2.80e-01 2.08e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06e-16 2.60e-08 0.00 1.00000
## FirstAuthorFemale1 3.09e-01 1.57e-01 1.97 0.05663 .
## Year1999 2.50e-03 3.96e-02 0.06 0.94995
## Year2000 2.51e-01 1.57e-01 1.60 0.11915
## Year2002 5.98e-01 8.76e-02 6.82 5.6e-08 ***
## Year2003 6.35e-01 3.17e-01 2.00 0.05301 .
## Year2004 3.30e-01 2.15e-01 1.54 0.13346
## Year2005 3.93e-01 1.52e-01 2.59 0.01387 *
## Year2006 3.40e-01 3.64e-01 0.93 0.35667
## Year2007 8.08e-01 9.60e-02 8.42 5.0e-10 ***
## Year2008 2.98e-01 1.95e-01 1.52 0.13643
## Year2009 2.52e-01 2.22e-01 1.14 0.26218
## Year2010 2.44e-01 1.82e-01 1.34 0.18913
## Year2011 1.35e+00 3.29e-01 4.11 0.00022 ***
## Year2012 7.54e-01 2.11e-01 3.58 0.00101 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared: 0.346, Adjusted R-squared: 0.0912
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2 is an outlier with |weight| <= 0.0012 ( < 0.002);
## 6 weights are ~= 1. The remaining 44 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.167 0.931 0.961 0.908 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 1.96e-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 4.972 1           2.230
## Year             4.972 13           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
## -7.19e-01 -2.78e-01  1.00e-15  2.36e-01  2.00e+00
##
## Coefficients:
```

```

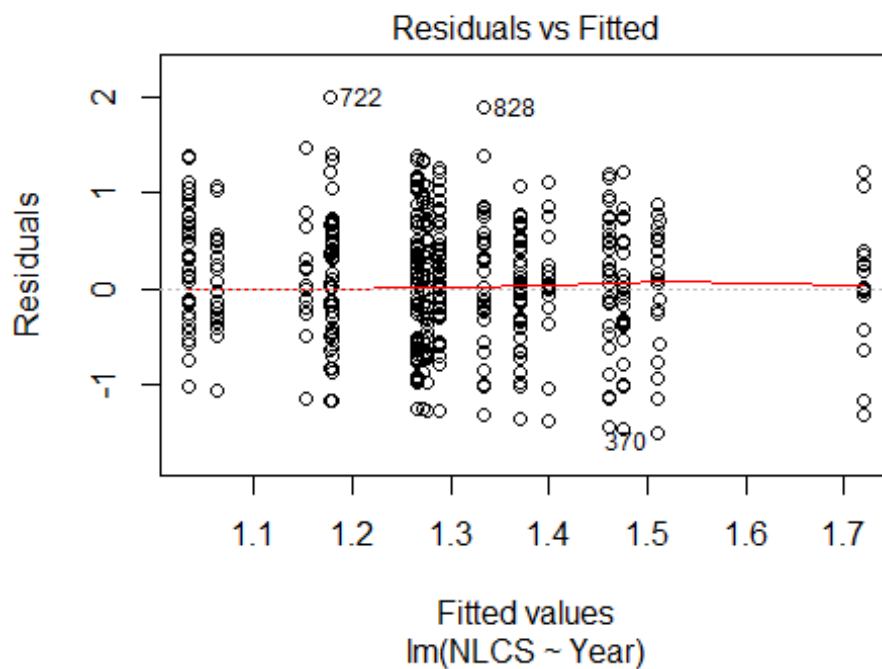
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    -1.00e-15   1.14e-08  0.00e+00   1.0000
## LastAuthorFemale1 2.05e-01   1.37e-01  1.49e+00   0.1439
## Year1999        8.72e-02   6.13e-01  1.40e-01   0.8877
## Year2000        5.60e-01   1.82e-08  3.07e+07  < 2e-16 ***
## Year2002        6.50e-01   1.03e-01  6.33e+00  2.5e-07 ***
## Year2003        7.19e-01   4.37e-01  1.65e+00   0.1086
## Year2004        3.79e-01   1.94e-01  1.96e+00   0.0582 .
## Year2005        3.78e-01   1.45e-01  2.60e+00   0.0135 *
## Year2006        2.94e-01   4.20e-01  7.00e-01   0.4874
## Year2007        8.08e-01   9.58e-02  8.44e+00  4.7e-10 ***
## Year2008        3.30e-01   1.93e-01  1.71e+00   0.0961 .
## Year2009        2.78e-01   2.32e-01  1.20e+00   0.2399
## Year2010        2.44e-01   1.81e-01  1.35e+00   0.1864
## Year2011        1.15e+00   3.45e-01  3.33e+00   0.0020 **
## Year2012        6.78e-01   2.03e-01  3.35e+00   0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.304, Adjusted R-squared:  0.0339
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 44 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0435 0.9220 0.9680 0.8940 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           1.96e-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: 51"
## [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
##   79   76   93   14   96   81   59   70   37   79   67   90   86  110  126
## 2011 2012
##  118  137
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   16   20    3   17   16   23   32   16   36   29   35   43   47   74
## 2011 2012
##   50   66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   20   16   17    1   14   14   19   25   14   31   23   30   34   41   51
## 2011 2012
##   41   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 = 14, df = 16, p-value = 0.6

```

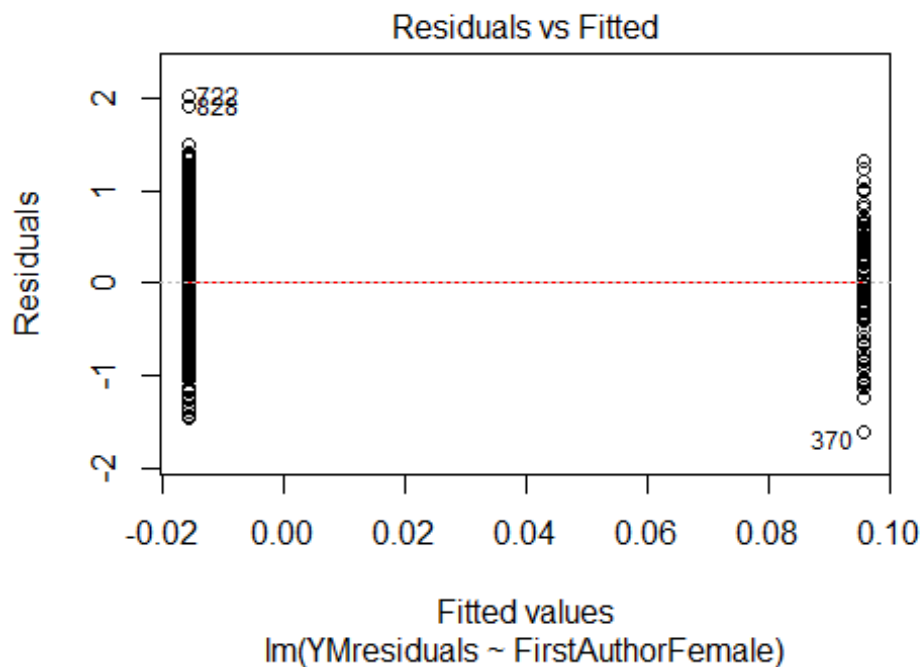


```

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

```

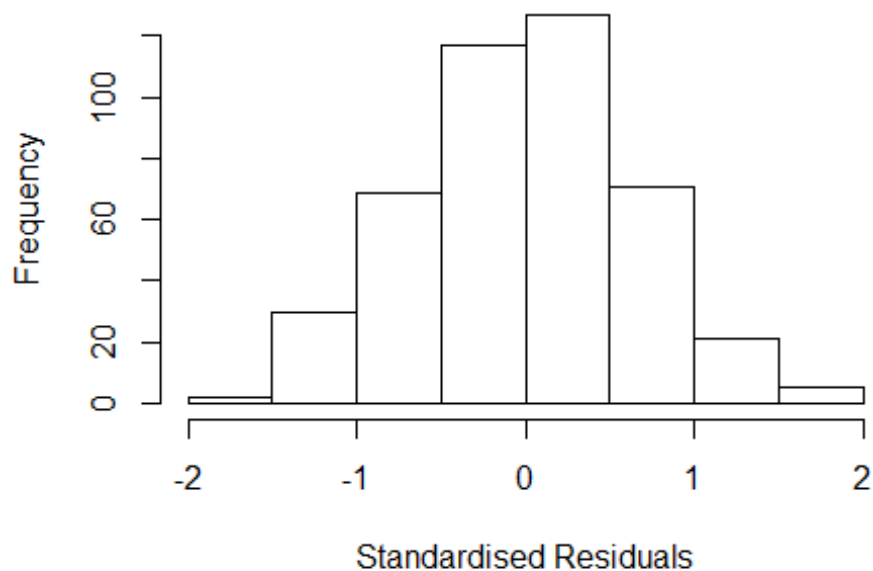
```
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.0296635898134"
## [1] "Male first author team size 2018 geometric mean: 2.85231848794071"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 380, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.03841247042755"
## [1] "Male last author team size 2018 geometric mean: 2.82271499887924"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 330, 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 2.154 1          1.468
## LastAuthorFemale  1.894 1          1.376
```

## UniqueAuthors	11.844	4	1.362
## Year	16.359	16	1.091

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
## -1.9661 -0.4378 0.0104 0.4414 1.9119
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.84241 0.12718 6.62 1.1e-10 ***
## FirstAuthorFemale1 0.06648 0.10537 0.63 0.5284
## LastAuthorFemale1 -0.10061 0.14064 -0.72 0.4748
## UniqueAuthors2 0.51059 0.10360 4.93 1.2e-06 ***
## UniqueAuthors3 0.66000 0.10393 6.35 5.6e-10 ***
## UniqueAuthors4 0.59948 0.09773 6.13 2.0e-09 ***
## UniqueAuthors5 0.35251 0.11379 3.10 0.0021 **
```

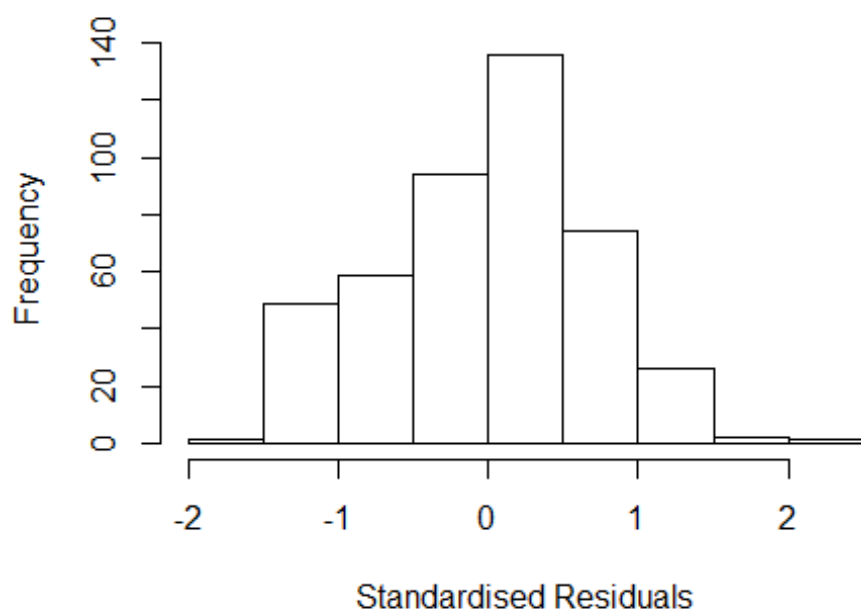


```

## Year1997          0.04953    0.21433    0.23    0.8174
## Year1998          0.09091    0.18962    0.48    0.6319
## Year1999          0.71259    0.14003    5.09    5.4e-07 ***
## Year2000          0.39723    0.26765    1.48    0.1385
## Year2001          0.43048    0.25148    1.71    0.0877 .
## Year2002         -0.00488    0.24946   -0.02    0.9844
## Year2003          0.17431    0.21211    0.82    0.4117
## Year2004         -0.05874    0.23403   -0.25    0.8019
## Year2005         -0.03686    0.17028   -0.22    0.8287
## Year2006          0.01430    0.19249    0.07    0.9408
## Year2007          0.11790    0.15433    0.76    0.4453
## Year2008         -0.24347    0.16844   -1.45    0.1491
## Year2009         -0.08118    0.16083   -0.50    0.6140
## Year2010         -0.12049    0.14654   -0.82    0.4114
## Year2011          0.09461    0.15409    0.61    0.5396
## Year2012         -0.04090    0.15785   -0.26    0.7957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.648
## Multiple R-squared:  0.185, Adjusted R-squared:  0.142
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 405 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.337  0.865  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      2.26e-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 1.822 1      1.350
## LastAuthorFemale  1.519 1      1.233
## Year              1.688 16      1.016

```

Residuals from first and last author



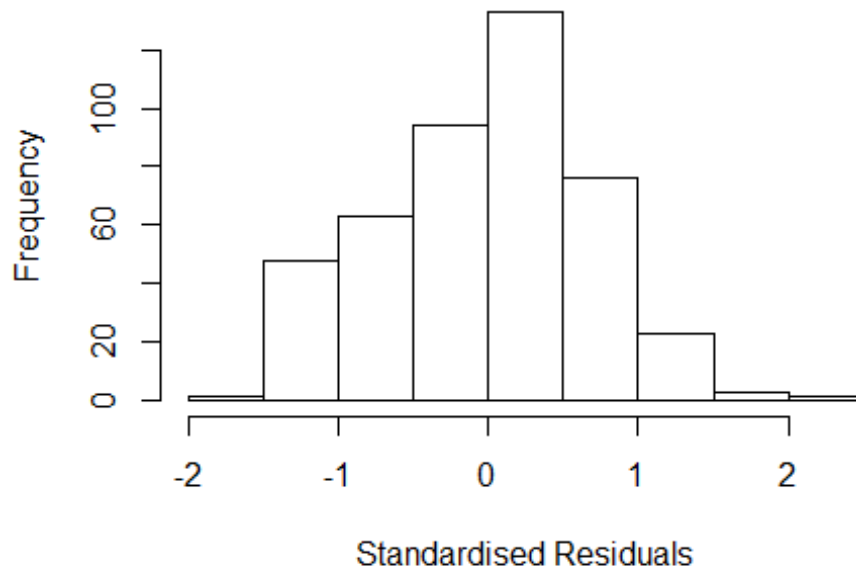
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6976 -0.4868 0.0541 0.4842 2.1035
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1002 0.1362 8.08 6.8e-15 ***
## FirstAuthorFemale1 0.1298 0.0995 1.30 0.193
## LastAuthorFemale1 -0.2242 0.1403 -1.60 0.111
## Year1997 0.0768 0.2095 0.37 0.714
## Year1998 0.2054 0.1994 1.03 0.304
## Year1999 1.1148 0.1362 8.19 3.2e-15 ***
## Year2000 0.4676 0.2694 1.74 0.083 .
## Year2001 0.5874 0.2337 2.51 0.012 *
## Year2002 0.1050 0.2384 0.44 0.660
## Year2003 0.2712 0.2375 1.14 0.254
## Year2004 -0.0247 0.2662 -0.09 0.926
## Year2005 0.1507 0.2030 0.74 0.458
```

```

## Year2006          0.3024      0.2051      1.47      0.141
## Year2007          0.1981      0.1851      1.07      0.285
## Year2008         -0.2015      0.1987     -1.01      0.311
## Year2009          0.0851      0.1832      0.46      0.642
## Year2010          0.0354      0.1627      0.22      0.828
## Year2011          0.2137      0.1644      1.30      0.194
## Year2012          0.1577      0.1742      0.91      0.366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.06,   Adjusted R-squared:  0.02
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 400 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.363  0.860  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      2.26e-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 1.435 1      1.198
## Year              1.435 16      1.011

```

Residuals from first author

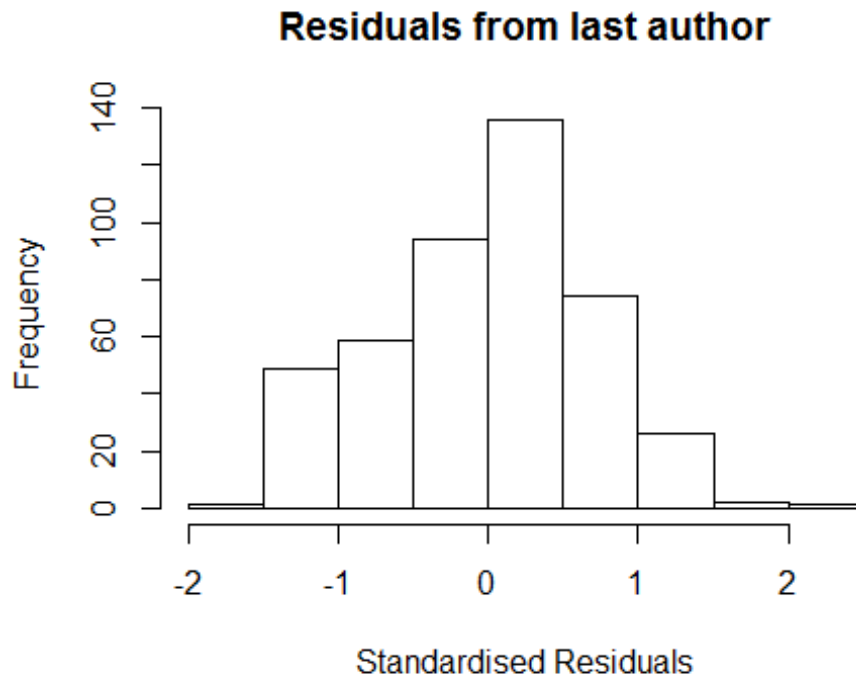


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6442 -0.5094 0.0505 0.4905 2.1253
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0944 0.1324 8.27 1.8e-15 ***
## FirstAuthorFemale1 0.0688 0.0930 0.74 0.4602
## Year1997 0.0334 0.2020 0.17 0.8688
## Year1998 0.2201 0.1964 1.12 0.2632
## Year1999 1.1206 0.1324 8.46 4.3e-16 ***
## Year2000 0.4810 0.2637 1.82 0.0689 .
## Year2001 0.6043 0.2314 2.61 0.0093 **
## Year2002 0.1081 0.2440 0.44 0.6579
## Year2003 0.2480 0.2349 1.06 0.2917
## Year2004 -0.0407 0.2744 -0.15 0.8822
## Year2005 0.1401 0.1999 0.70 0.4836
## Year2006 0.2985 0.2034 1.47 0.1429
```

```

## Year2007          0.1809      0.1800      1.00      0.3156
## Year2008         -0.2007      0.1946     -1.03      0.3030
## Year2009          0.0817      0.1811      0.45      0.6519
## Year2010          0.0341      0.1596      0.21      0.8310
## Year2011          0.2139      0.1626      1.32      0.1891
## Year2012          0.1560      0.1719      0.91      0.3649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.054, Adjusted R-squared:  0.0161
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 398 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.352  0.861  0.950  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      2.26e-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 1.233 1          1.110
## Year            1.233 16          1.007

```



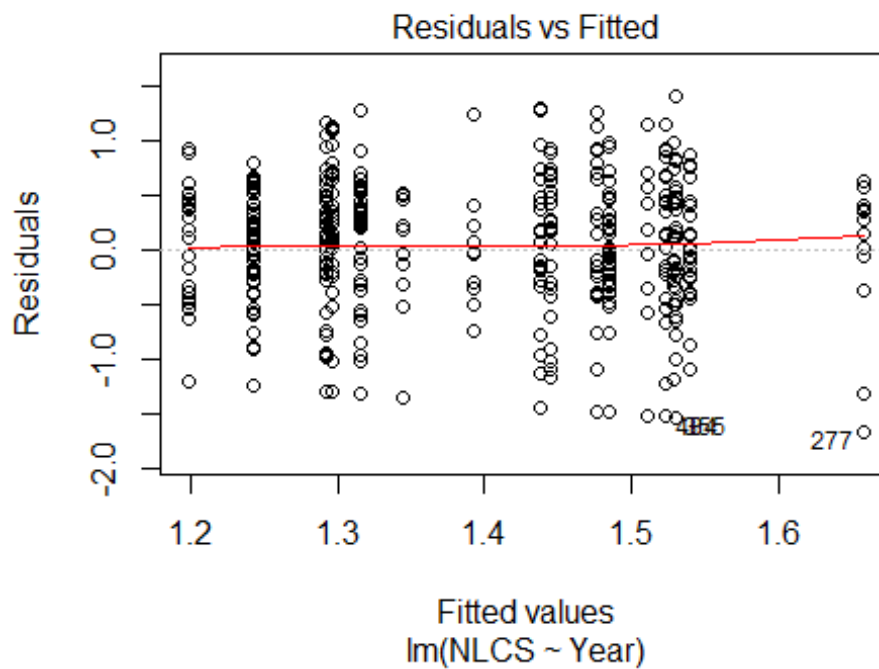
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5849 -0.4921 0.0626 0.4760 2.0788
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1117 0.1321 8.42 6.0e-16 ***
## LastAuthorFemale1 -0.1665 0.1264 -1.32 0.189
## Year1997 0.0745 0.2080 0.36 0.720
## Year1998 0.2128 0.1969 1.08 0.280
## Year1999 1.1033 0.1321 8.35 9.6e-16 ***
## Year2000 0.4731 0.2600 1.82 0.069 .
## Year2001 0.5994 0.2323 2.58 0.010 *
## Year2002 0.1301 0.2391 0.54 0.587
## Year2003 0.2773 0.2355 1.18 0.240
## Year2004 -0.0115 0.2640 -0.04 0.965
## Year2005 0.1624 0.2013 0.81 0.420
## Year2006 0.2930 0.2042 1.43 0.152
```

```

## Year2007          0.1855      0.1813      1.02      0.307
## Year2008          -0.2000     0.1945     -1.03     0.304
## Year2009          0.0883      0.1814      0.49     0.627
## Year2010          0.0333      0.1600      0.21     0.836
## Year2011          0.2236      0.1618      1.38     0.168
## Year2012          0.1580      0.1710      0.92     0.356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.0567, Adjusted R-squared:  0.0188
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.371  0.855  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      2.26e-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: 442"
## [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
##   57   35   55   63   61   71   57   53   24   62   46   76   67   83   85
## 2011 2012
##  100   97
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   10   16   10   13   23   25   25   12   26   23   30   36   38   46
## 2011 2012

```

```
## 49 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 19 9 15 8 12 17 21 19 9 20 17 25 32 33 32
## 2011 2012
## 41 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 = 21, df = 16, p-value = 0.2
```



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

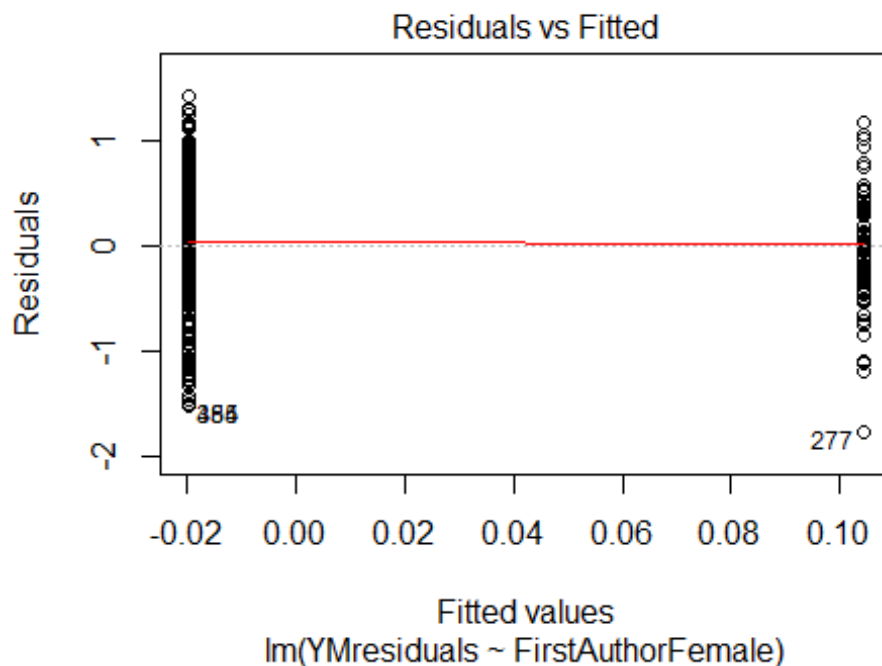
## [1] "Female first author team size 2018 geometric mean: 2.16093273781394"
## [1] "Male first author team size 2018 geometric mean: 3.26425567626842"

## 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 = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.58619041845698"
## [1] "Male last author team size 2018 geometric mean: 3.14125158935718"

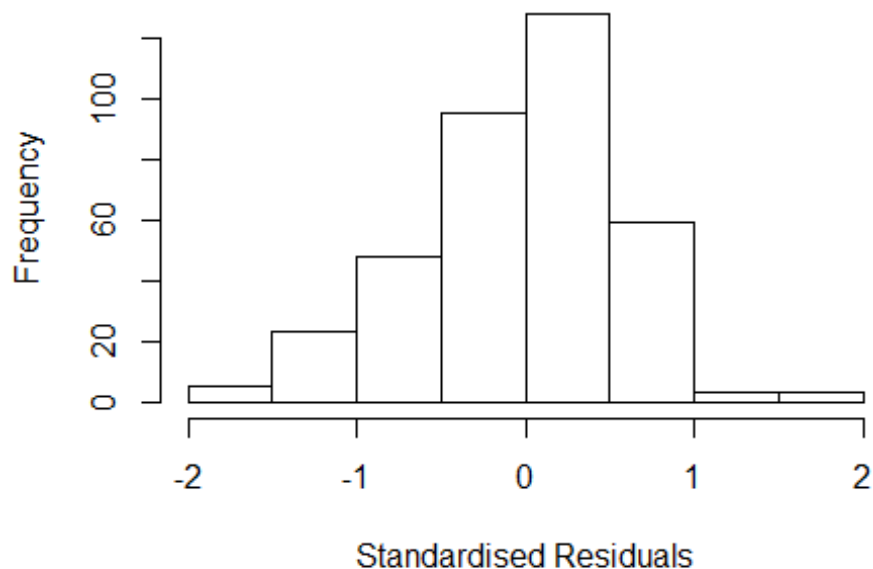
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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	1.322	1	1.150
LastAuthorFemale	1.858	1	1.363
UniqueAuthors	2.813	4	1.138
Year	4.913	16	1.051

Residuals from first and last author and team size



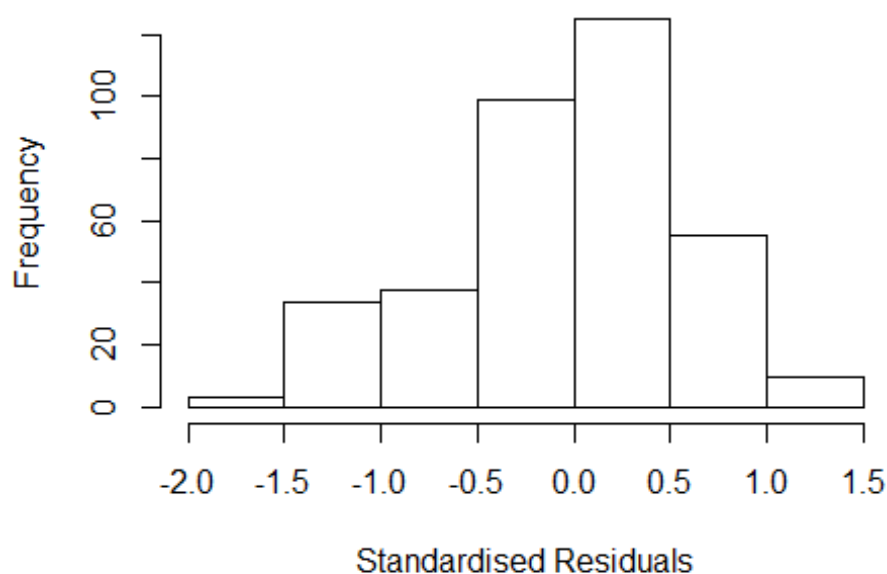
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.9971 -0.3768 0.0455 0.3920 1.8046
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.945550 0.134960 7.01 1.3e-11 ***
## FirstAuthorFemale1 0.020797 0.080577 0.26 0.79649
## LastAuthorFemale1 -0.069166 0.098962 -0.70 0.48508
## UniqueAuthors2 0.465214 0.121490 3.83 0.00015 ***
## UniqueAuthors3 0.563905 0.113634 4.96 1.1e-06 ***
## UniqueAuthors4 0.480869 0.115057 4.18 3.7e-05 ***
## UniqueAuthors5 0.297121 0.120708 2.46 0.01433 *
## Year1997 0.189740 0.210530 0.90 0.36809
## Year1998 0.215757 0.168782 1.28 0.20201
## Year1999 0.082707 0.256457 0.32 0.74727
```

```

## Year2000      0.466866  0.199297  2.34  0.01973 *
## Year2001      0.182851  0.210577  0.87  0.38582
## Year2002      0.008469  0.241905  0.04  0.97209
## Year2003      0.113039  0.224113  0.50  0.61432
## Year2004     -0.089807  0.187651 -0.48  0.63254
## Year2005      0.003886  0.174758  0.02  0.98227
## Year2006      0.024208  0.160133  0.15  0.87993
## Year2007      0.154275  0.172508  0.89  0.37179
## Year2008      0.017375  0.174656  0.10  0.92081
## Year2009      0.000815  0.162564  0.01  0.99601
## Year2010     -0.032618  0.170878 -0.19  0.84873
## Year2011     -0.055629  0.148421 -0.37  0.70804
## Year2012      0.128353  0.143385  0.90  0.37133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.149, Adjusted R-squared:  0.0938
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 337 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.162  0.866  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      2.75e-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 1.209 1      1.100
## LastAuthorFemale  1.610 1      1.269
## Year              1.792 16      1.018

```

Residuals from first and last author



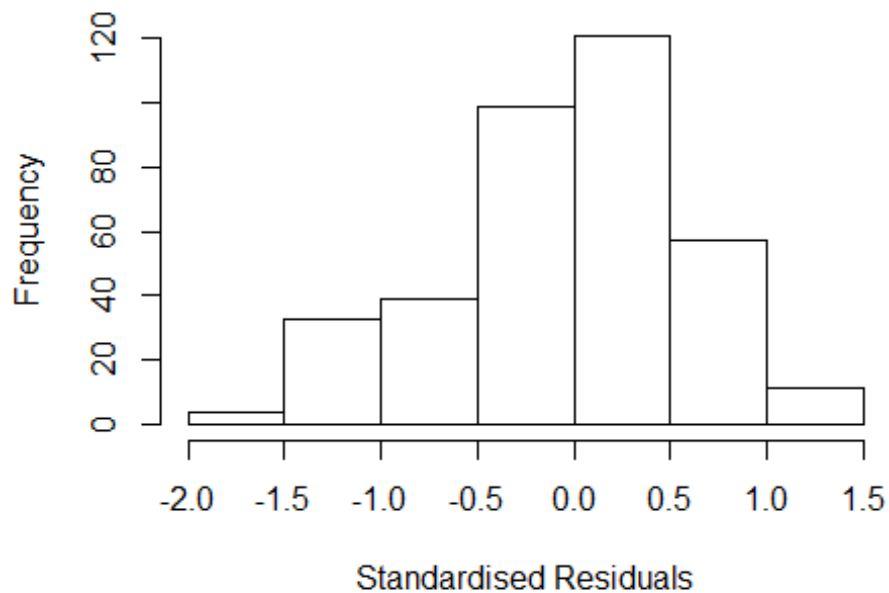
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8730 -0.4168 0.0573 0.4005 1.4206
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22681 0.13560 9.05 <2e-16 ***
## FirstAuthorFemale1 0.06564 0.08420 0.78 0.4361
## LastAuthorFemale1 -0.07177 0.11091 -0.65 0.5180
## Year1997 0.17084 0.22561 0.76 0.4494
## Year1998 0.28619 0.18798 1.52 0.1288
## Year1999 0.29663 0.27836 1.07 0.2873
## Year2000 0.58053 0.21593 2.69 0.0075 **
## Year2001 0.28555 0.20230 1.41 0.1590
## Year2002 0.16530 0.23965 0.69 0.4908
## Year2003 0.23785 0.24991 0.95 0.3419
## Year2004 0.04695 0.21561 0.22 0.8277
## Year2005 0.10511 0.21171 0.50 0.6199
```

```

## Year2006          0.19823    0.17150    1.16    0.2485
## Year2007          0.28086    0.18686    1.50    0.1337
## Year2008          0.10305    0.20360    0.51    0.6131
## Year2009          0.11590    0.18350    0.63    0.5280
## Year2010         -0.01870    0.18021   -0.10    0.9174
## Year2011          0.00705    0.16195    0.04    0.9653
## Year2012          0.29492    0.15776    1.87    0.0624 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0479, Adjusted R-squared:  -0.00175
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 339 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.282  0.868  0.950  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      2.75e-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 1.171 1      1.082
## Year              1.171 16      1.005

```

Residuals from first author



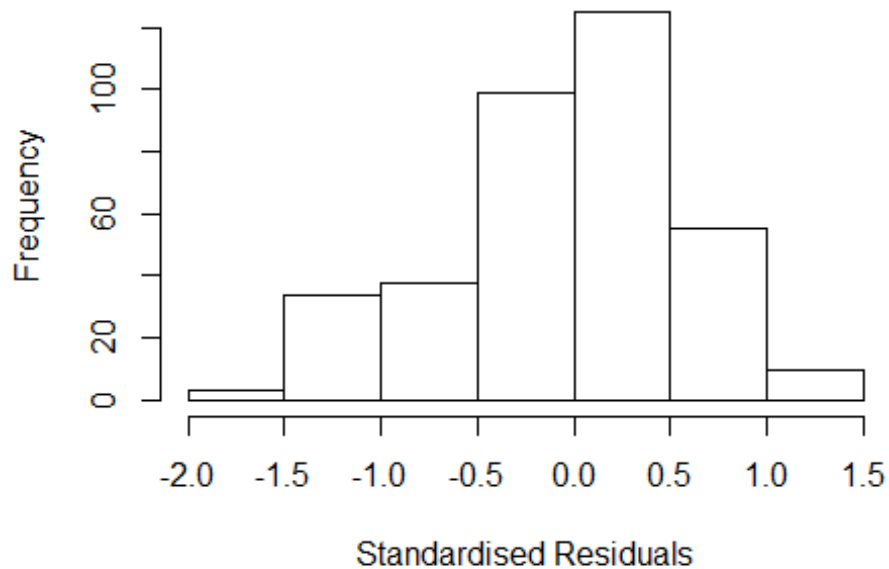
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8636 -0.4046 0.0672 0.4113 1.4218
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22324 0.13431 9.11 <2e-16 ***
## FirstAuthorFemale1 0.05492 0.08371 0.66 0.5123
## Year1997 0.14325 0.22300 0.64 0.5211
## Year1998 0.29011 0.18500 1.57 0.1178
## Year1999 0.29031 0.27858 1.04 0.2981
## Year2000 0.58546 0.21619 2.71 0.0071 **
## Year2001 0.28795 0.20290 1.42 0.1568
## Year2002 0.16016 0.23992 0.67 0.5049
## Year2003 0.22446 0.24721 0.91 0.3645
## Year2004 0.05313 0.21543 0.25 0.8054
## Year2005 0.09907 0.20972 0.47 0.6369
## Year2006 0.19893 0.16972 1.17 0.2420
```

```

## Year2007          0.26730      0.18295      1.46      0.1449
## Year2008          0.09524      0.20096      0.47      0.6359
## Year2009          0.11553      0.18393      0.63      0.5303
## Year2010         -0.01513      0.17992     -0.08      0.9330
## Year2011          0.00994      0.16141      0.06      0.9510
## Year2012          0.29367      0.15722      1.87      0.0626 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0468, Adjusted R-squared:  7.77e-06
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 341 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.287  0.869   0.950   0.896   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      2.75e-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 1.555 1      1.247
## Year              1.555 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
## -1.8181 -0.4006 0.0482 0.4099 1.4044
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23347 0.13285 9.29 <2e-16 ***
## LastAuthorFemale1 -0.05599 0.10645 -0.53 0.5992
## Year1997 0.17970 0.22347 0.80 0.4219
## Year1998 0.28998 0.18550 1.56 0.1189
## Year1999 0.30791 0.27548 1.12 0.2645
## Year2000 0.58459 0.21451 2.73 0.0068 **
## Year2001 0.29510 0.20174 1.46 0.1444
## Year2002 0.16735 0.23842 0.70 0.4832
## Year2003 0.24134 0.24793 0.97 0.3310
## Year2004 0.05672 0.21756 0.26 0.7945
## Year2005 0.11137 0.21256 0.52 0.6006
## Year2006 0.19522 0.16981 1.15 0.2511
```

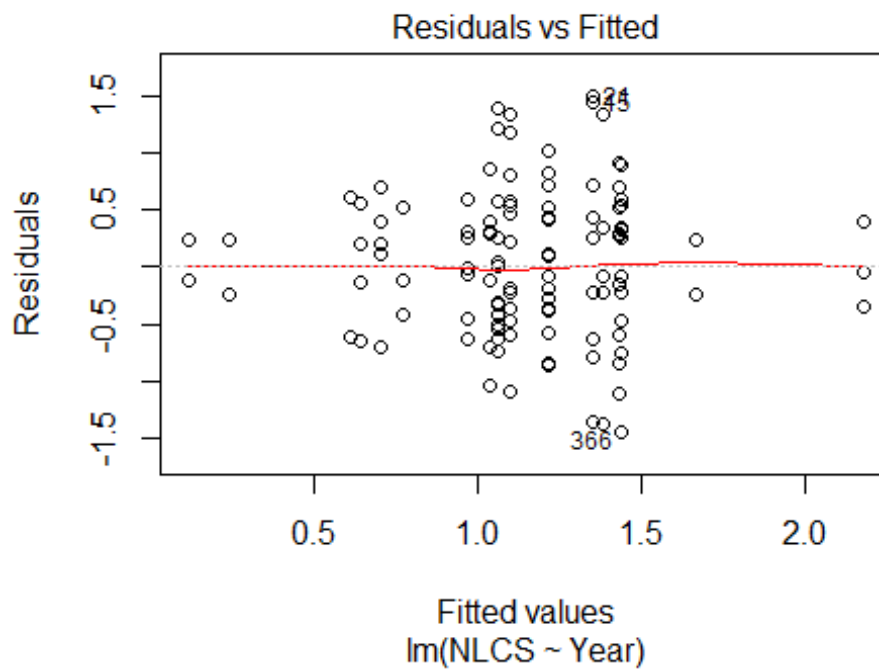


```

## Year2007      0.27721      0.18362      1.51      0.1320
## Year2008      0.10794      0.20079      0.54      0.5912
## Year2009      0.12566      0.18019      0.70      0.4860
## Year2010     -0.01830      0.17873     -0.10      0.9185
## Year2011      0.00932      0.16036      0.06      0.9537
## Year2012      0.29561      0.15481      1.91      0.0570 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.0466, Adjusted R-squared:  -0.000211
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 342 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.307  0.875   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      2.75e-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: 364"
## [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
##   26   11   15   15    7   10   11   28    7   15   17   19   35   18   16
## 2011 2012
##   29   23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10    3    3    4    2    3    2    9    2    5    7    6   15    7   13
## 2011 2012

```

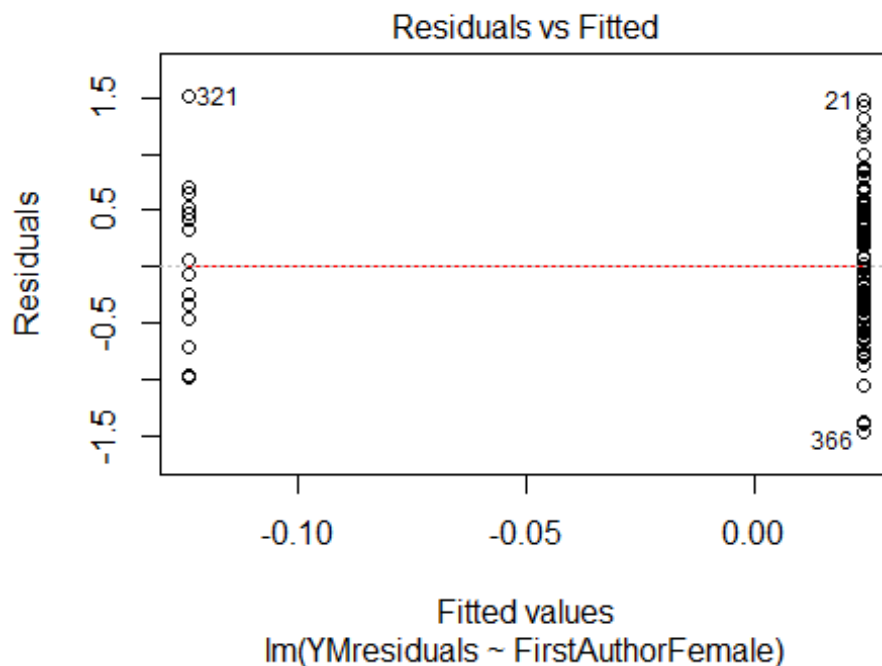
```
## 17 11
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 3 3 4 2 3 1 9 2 3 5 6 15 6 9
## 2011 2012
## 15 6
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.38, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 2.23385451109417"
## [1] "Male first author team size 2018 geometric mean: 2.09506136298666"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

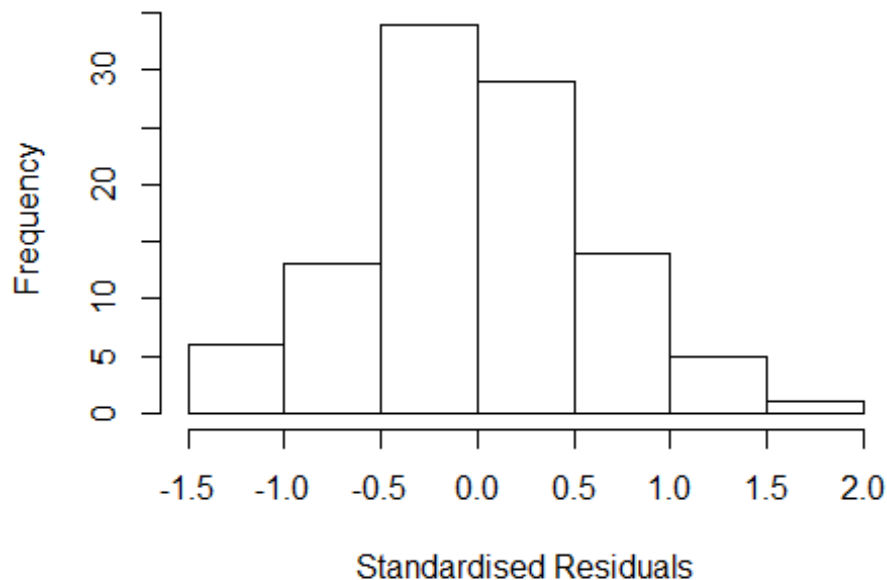
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.25869387091371"
## [1] "Male last author team size 2018 geometric mean: 2.09043383528137"

## 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 = 0.8
## 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  6.025  1      2.455
## LastAuthorFemale  3.051  1      1.747
## UniqueAuthors    311.964  4      2.050
## Year              484.189 16      1.213
```

Residuals from first and last author and team size



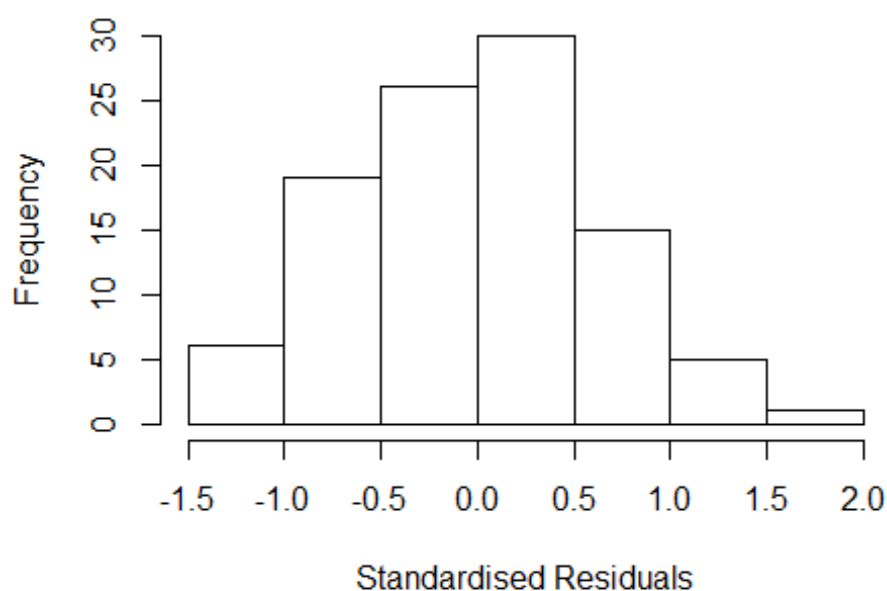
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4987 -0.3618 -0.0217  0.3076  1.5044
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2926    0.3882   3.33  0.0013 **
## FirstAuthorFemale1  0.0549    0.2828   0.19  0.8467
## LastAuthorFemale1 -0.4086    0.2642  -1.55  0.1260
## UniqueAuthors2     0.2124    0.2071   1.03  0.3083
## UniqueAuthors3     0.6858    0.2288   3.00  0.0036 **
## UniqueAuthors4     0.4351    0.2444   1.78  0.0788 .
## UniqueAuthors5     0.6619    0.1918   3.45  0.0009 ***
## Year1997          -0.9616    0.5209  -1.85  0.0686 .
## Year1998          -1.2482    0.4182  -2.98  0.0038 **
## Year1999          -0.7022    0.4723  -1.49  0.1411
```

```

## Year2000          -1.1598      0.4151    -2.79    0.0065 **
## Year2001           0.6823      0.5453     1.25    0.2145
## Year2002          -1.9784      0.4530    -4.37    3.8e-05 ***
## Year2003          -0.0113      0.4604    -0.02    0.9804
## Year2004           0.3754      0.4264     0.88    0.3813
## Year2005          -0.2996      0.6014    -0.50    0.6198
## Year2006          -0.8458      0.4518    -1.87    0.0649 .
## Year2007          -0.9139      0.5224    -1.75    0.0841 .
## Year2008          -0.2839      0.4573    -0.62    0.5364
## Year2009          -0.6041      0.5278    -1.14    0.2559
## Year2010          -0.5807      0.4721    -1.23    0.2224
## Year2011          -0.3473      0.4406    -0.79    0.4330
## Year2012          -0.4636      0.4373    -1.06    0.2922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.367, Adjusted R-squared:  0.191
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 89 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.566  0.888  0.962   0.915   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      9.80e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.261 1      1.806
## LastAuthorFemale  2.324 1      1.524
## Year              2.601 16      1.030

```

Residuals from first and last author

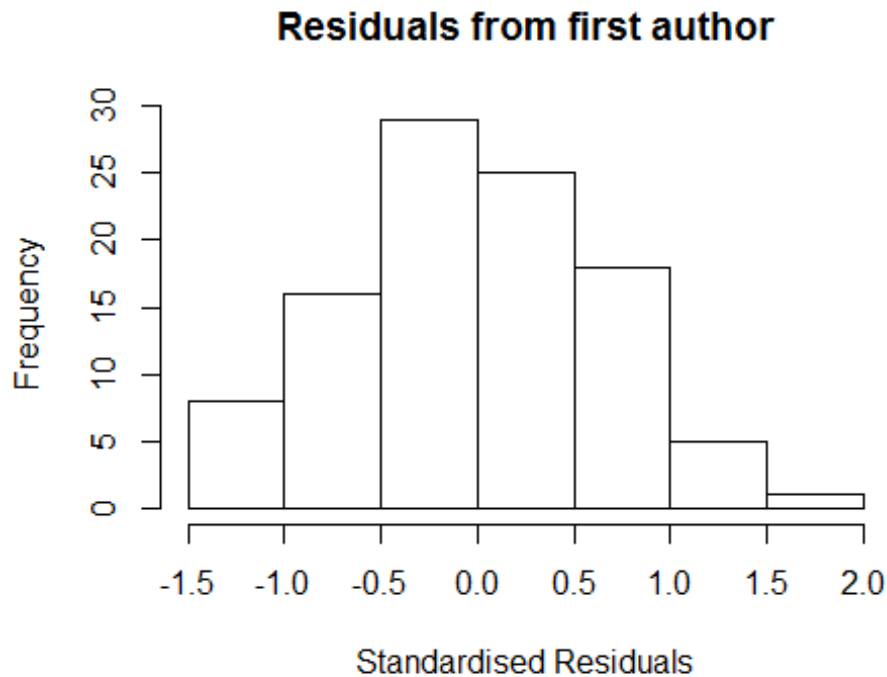


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3300 -0.4329 0.0113 0.4233 1.5240
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33000 0.41170 3.23 0.0018 **
## FirstAuthorFemale1 0.15232 0.24804 0.61 0.5408
## LastAuthorFemale1 -0.62955 0.26780 -2.35 0.0211 *
## Year1997 -0.56233 0.47343 -1.19 0.2383
## Year1998 -1.21452 0.42253 -2.87 0.0051 **
## Year1999 -0.68147 0.47337 -1.44 0.1537
## Year2000 -1.09100 0.44629 -2.44 0.0166 *
## Year2001 0.84824 0.44883 1.89 0.0623 .
## Year2002 -1.33000 0.41170 -3.23 0.0018 **
## Year2003 0.26956 0.45987 0.59 0.5593
## Year2004 0.33800 0.44745 0.76 0.4522
## Year2005 -0.27761 0.64372 -0.43 0.6674
```

```

## Year2006          -0.46971      0.44149      -1.06      0.2905
## Year2007          -0.67404      0.48050      -1.40      0.1644
## Year2008          -0.17874      0.47046      -0.38      0.7050
## Year2009          -0.38271      0.50247      -0.76      0.4484
## Year2010          -0.24661      0.47734      -0.52      0.6068
## Year2011          -0.09897      0.44083      -0.22      0.8229
## Year2012           0.00151      0.46999       0.00      0.9974
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.743
## Multiple R-squared:  0.283, Adjusted R-squared:  0.128
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 98 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.653  0.913  0.968  0.935  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      9.80e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.862 1          1.365
## Year              1.862 16          1.020

```



```
## [1] "List of 0 outliers 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
## -1.3316 -0.4832 -0.0238 0.4733 1.5224
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.332 0.406 3.28 0.0015 **
## FirstAuthorFemale1 -0.101 0.207 -0.49 0.6264
## Year1997 -0.564 0.468 -1.20 0.2319
## Year1998 -1.216 0.417 -2.92 0.0045 **
## Year1999 -0.683 0.468 -1.46 0.1481
## Year2000 -1.093 0.441 -2.48 0.0152 *
## Year2001 0.847 0.443 1.91 0.0596 .
## Year2002 -1.332 0.406 -3.28 0.0015 **
## Year2003 0.156 0.484 0.32 0.7478
## Year2004 0.336 0.442 0.76 0.4488
## Year2005 -0.282 0.638 -0.44 0.6601
## Year2006 -0.497 0.441 -1.13 0.2629
```

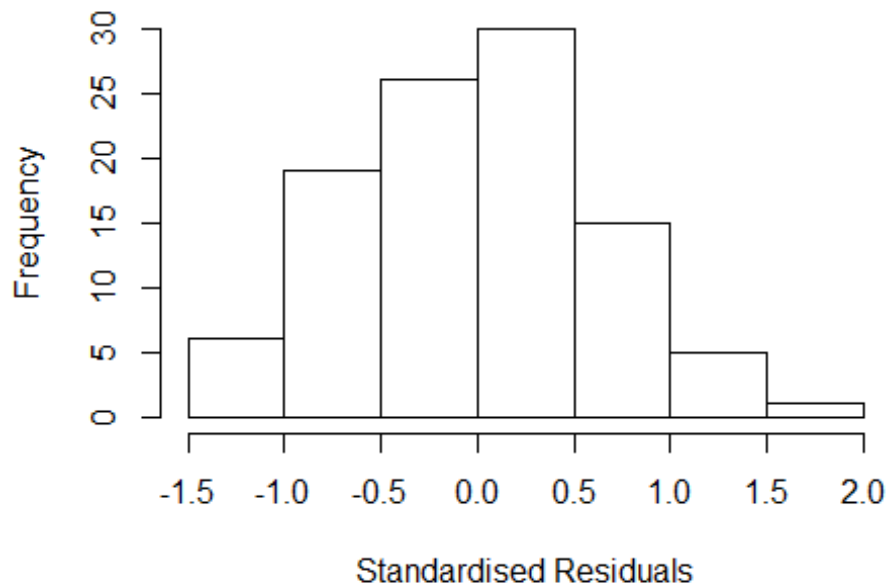


```

## Year2007          -0.583      0.485   -1.20   0.2331
## Year2008          -0.218      0.459   -0.48   0.6357
## Year2009          -0.336      0.507   -0.66   0.5096
## Year2010          -0.204      0.487   -0.42   0.6770
## Year2011          -0.164      0.432   -0.38   0.7054
## Year2012          -0.143      0.498   -0.29   0.7753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.76
## Multiple R-squared:  0.241, Adjusted R-squared:  0.087
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 96 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.668  0.913  0.961  0.933  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      9.80e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.357 1      1.165
## Year              1.357 16      1.010

```

Residuals from last author



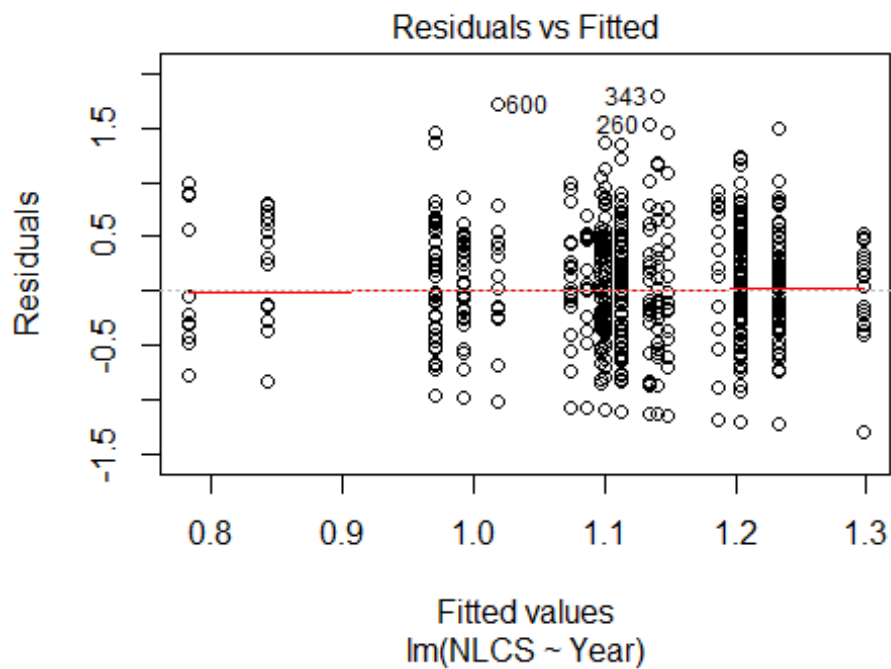
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33183 -0.48895 0.00156 0.43553 1.52217
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3318 0.4050 3.29 0.0015 **
## LastAuthorFemale1 -0.5355 0.2090 -2.56 0.0122 *
## Year1997 -0.5639 0.4673 -1.21 0.2309
## Year1998 -1.2163 0.4159 -2.92 0.0044 **
## Year1999 -0.6835 0.4673 -1.46 0.1473
## Year2000 -1.0928 0.4400 -2.48 0.0150 *
## Year2001 0.8465 0.4426 1.91 0.0592 .
## Year2002 -1.3318 0.4050 -3.29 0.0015 **
## Year2003 0.2871 0.4572 0.63 0.5318
## Year2004 0.3362 0.4412 0.76 0.4482
## Year2005 -0.2822 0.6369 -0.44 0.6589
## Year2006 -0.4279 0.4298 -1.00 0.3222
```

```

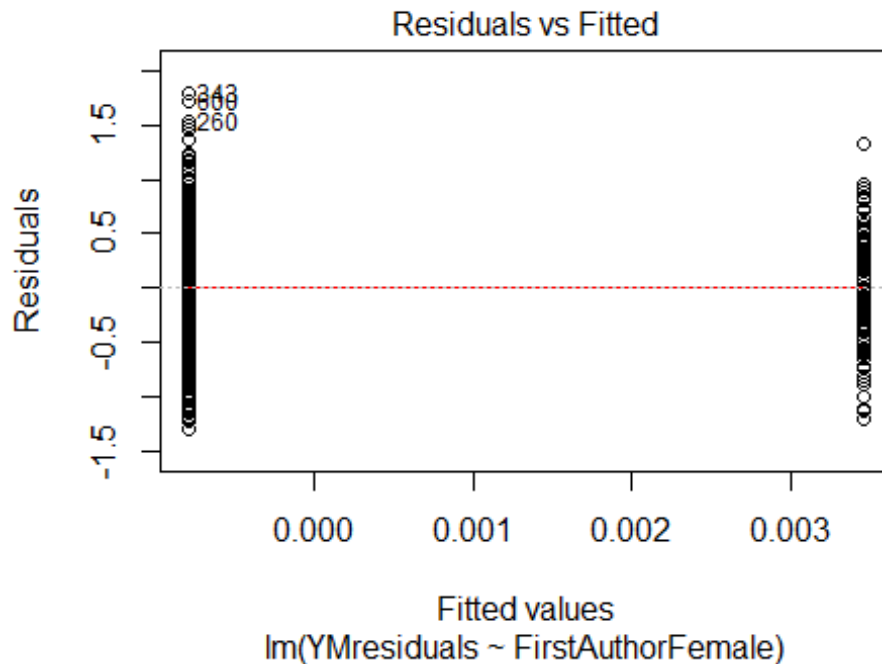
## Year2007          -0.6201      0.4682    -1.32    0.1890
## Year2008          -0.1623      0.4614    -0.35    0.7258
## Year2009          -0.3551      0.4981    -0.71    0.4779
## Year2010          -0.2192      0.4822    -0.45    0.6505
## Year2011          -0.0820      0.4337    -0.19    0.8505
## Year2012          -0.0018      0.4727     0.00    0.9970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.762
## Multiple R-squared:  0.277, Adjusted R-squared:  0.13
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 94 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.670  0.917  0.961  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      9.80e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 102"
## [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
##   59   48   42   59   62   66   51   43   54   45   70  100   81  117  155
## 2011 2012
##  160  193
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   18   12   19   26   28   17   17   23   18   35   47   46   61   81
## 2011 2012

```

```
##      82   111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   15   11   18   24   23   15   16   20   17   34   35   39   53   66
## 2011 2012
##   72   94
## [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
```

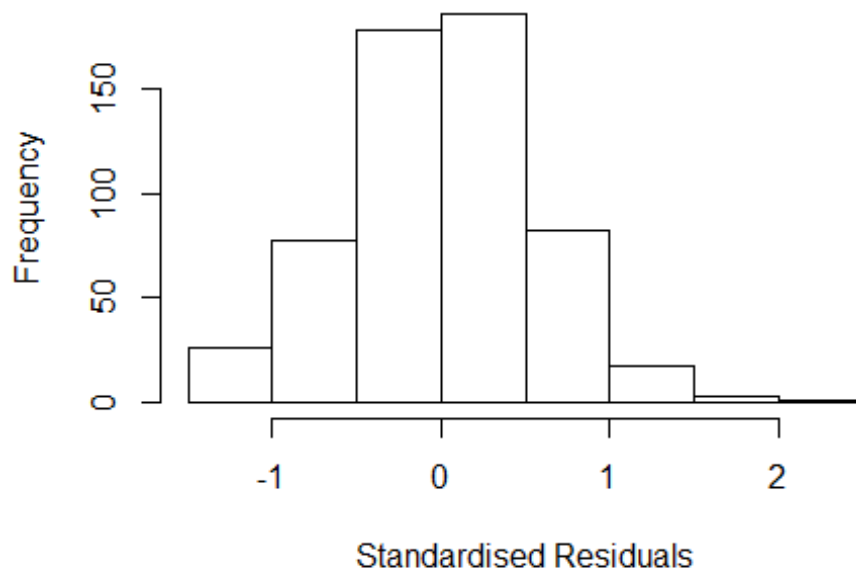


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.8, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 2.92982467052656"
## [1] "Male first author team size 2018 geometric mean: 3.16946844325283"
##
## 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: 2.79861607129972"
## [1] "Male last author team size 2018 geometric mean: 3.18877856991098"
##
## 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 1.574 1      1.255
## LastAuthorFemale  1.704 1      1.305
## UniqueAuthors    1.688 4      1.068
## Year              1.992 16     1.022
```

Residuals from first and last author and team size



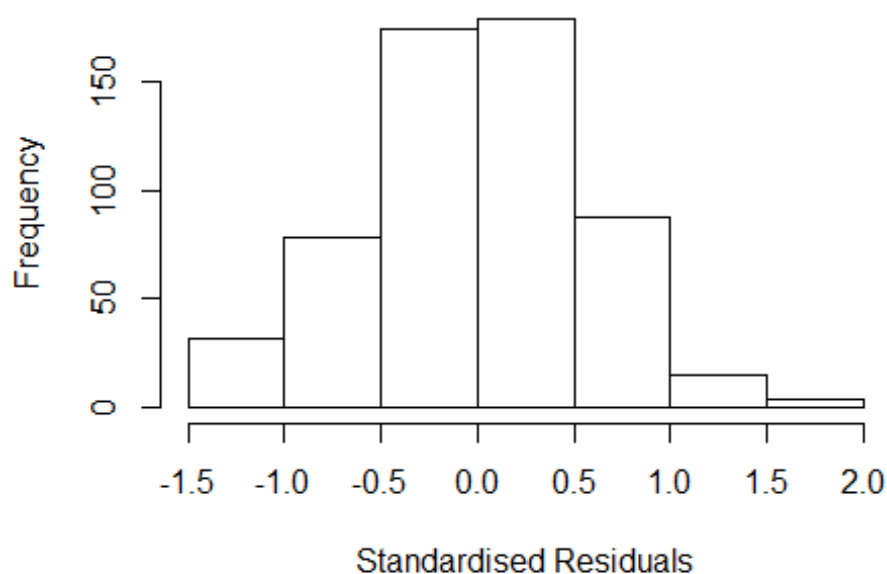
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32348 -0.34308 0.00486 0.39445 2.05701
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7749 0.1843 4.21 3.0e-05 ***
## FirstAuthorFemale1 0.0512 0.0674 0.76 0.44770
## LastAuthorFemale1 -0.1373 0.0837 -1.64 0.10155
## UniqueAuthors2 0.1808 0.0636 2.84 0.00467 **
## UniqueAuthors3 0.2580 0.0750 3.44 0.00063 ***
## UniqueAuthors4 0.3733 0.0889 4.20 3.1e-05 ***
## UniqueAuthors5 0.2600 0.0821 3.17 0.00162 **
## Year1997 0.3230 0.2088 1.55 0.12238
## Year1998 -0.1845 0.2594 -0.71 0.47730
## Year1999 0.4012 0.2084 1.93 0.05470 .
```

```

## Year2000      0.1642      0.2386      0.69  0.49154
## Year2001      0.1011      0.2469      0.41  0.68252
## Year2002      0.3425      0.2875      1.19  0.23411
## Year2003      0.3677      0.3028      1.21  0.22508
## Year2004      0.1723      0.2121      0.81  0.41691
## Year2005      0.0516      0.2251      0.23  0.81872
## Year2006      0.1613      0.2080      0.78  0.43839
## Year2007      0.2377      0.1963      1.21  0.22647
## Year2008     -0.0474      0.2040     -0.23  0.81618
## Year2009      0.2230      0.2024      1.10  0.27086
## Year2010      0.2056      0.1988      1.03  0.30139
## Year2011      0.2727      0.1935      1.41  0.15929
## Year2012      0.2380      0.1936      1.23  0.21963
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0649
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 523 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.118  0.864  0.948  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      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 1.431 1      1.196
## LastAuthorFemale  1.420 1      1.192
## Year              1.295 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
## -1.32160 -0.36562 0.00179 0.41392 1.82533
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.85784 0.18391 4.66 3.9e-06 ***
## FirstAuthorFemale1 0.09429 0.06634 1.42 0.156
## LastAuthorFemale1 -0.19659 0.08136 -2.42 0.016 *
## Year1997 0.39436 0.21542 1.83 0.068 .
## Year1998 -0.17226 0.27292 -0.63 0.528
## Year1999 0.46376 0.20433 2.27 0.024 *
## Year2000 0.21504 0.24388 0.88 0.378
## Year2001 0.24984 0.24138 1.04 0.301
## Year2002 0.38079 0.27382 1.39 0.165
## Year2003 0.38032 0.30807 1.23 0.218
## Year2004 0.19653 0.21889 0.90 0.370
## Year2005 0.09842 0.23483 0.42 0.675
```

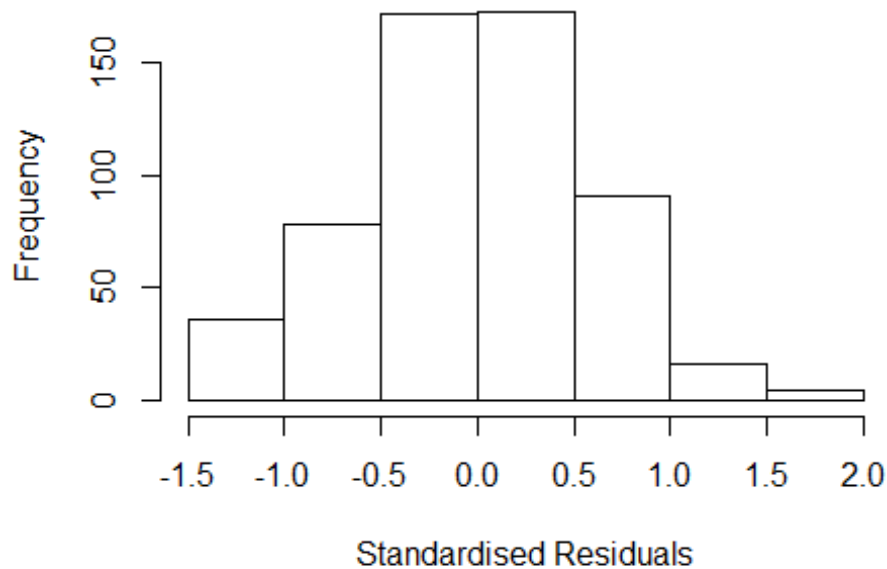


```

## Year2006          0.19399      0.20616      0.94      0.347
## Year2007          0.28426      0.19845      1.43      0.153
## Year2008          0.00822      0.20802      0.04      0.969
## Year2009          0.26803      0.20614      1.30      0.194
## Year2010          0.28825      0.20163      1.43      0.153
## Year2011          0.35569      0.19401      1.83      0.067 .
## Year2012          0.35027      0.19486      1.80      0.073 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0556, Adjusted R-squared:  0.0247
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 528 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.276  0.867   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      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 1.144 1      1.070
## Year              1.144 16      1.004

```

Residuals from first author



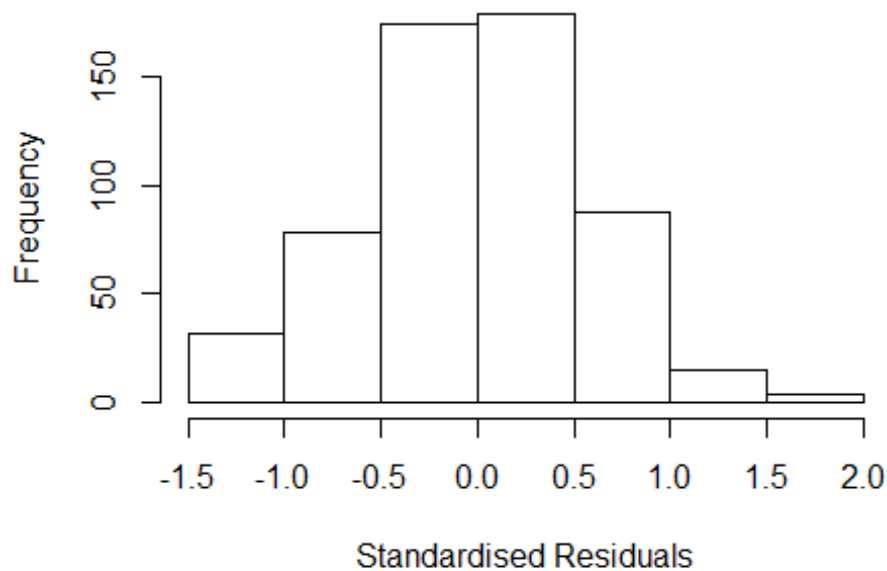
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3134 -0.3643 -0.0029 0.4076 1.8215
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8593 0.1805 4.76 2.5e-06 ***
## FirstAuthorFemale1 0.0196 0.0605 0.32 0.746
## Year1997 0.3682 0.2113 1.74 0.082 .
## Year1998 -0.2099 0.2675 -0.78 0.433
## Year1999 0.4541 0.2022 2.25 0.025 *
## Year2000 0.2344 0.2379 0.98 0.325
## Year2001 0.2522 0.2409 1.05 0.296
## Year2002 0.3456 0.2717 1.27 0.204
## Year2003 0.3511 0.2983 1.18 0.240
## Year2004 0.1882 0.2185 0.86 0.389
## Year2005 0.0729 0.2336 0.31 0.755
## Year2006 0.1598 0.2041 0.78 0.434
```

```

## Year2007          0.2810      0.1957      1.44      0.152
## Year2008          0.0133      0.2049      0.06      0.948
## Year2009          0.2671      0.2032      1.31      0.189
## Year2010          0.2582      0.1984      1.30      0.194
## Year2011          0.3386      0.1910      1.77      0.077 .
## Year2012          0.3338      0.1914      1.74      0.082 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.0452, Adjusted R-squared:  0.0158
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 531 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.292  0.874  0.952  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      1.75e-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 1.146 1      1.071
## Year              1.146 16      1.004

```

Residuals from last author



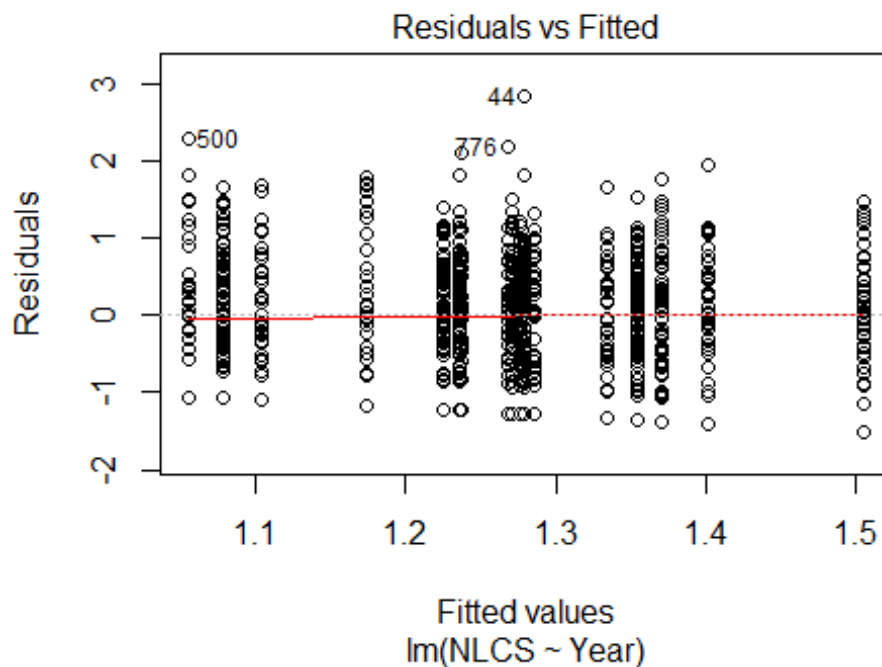
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32451 -0.35731 -0.00726 0.42272 1.82083
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8607 0.1814 4.75 2.7e-06 ***
## LastAuthorFemale1 -0.1486 0.0730 -2.04 0.042 *
## Year1997 0.3907 0.2138 1.83 0.068 .
## Year1998 -0.1756 0.2694 -0.65 0.515
## Year1999 0.4638 0.2017 2.30 0.022 *
## Year2000 0.2487 0.2356 1.06 0.292
## Year2001 0.2514 0.2407 1.04 0.297
## Year2002 0.3891 0.2676 1.45 0.146
## Year2003 0.3740 0.3076 1.22 0.225
## Year2004 0.2042 0.2153 0.95 0.343
## Year2005 0.1115 0.2345 0.48 0.635
## Year2006 0.2101 0.2030 1.03 0.301
```

```

## Year2007          0.2855      0.1963      1.45      0.146
## Year2008          0.0217      0.2053      0.11      0.916
## Year2009          0.2839      0.2024      1.40      0.161
## Year2010          0.2976      0.1990      1.50      0.135
## Year2011          0.3648      0.1912      1.91      0.057 .
## Year2012          0.3635      0.1917      1.90      0.058 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0525, Adjusted R-squared:  0.0233
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 525 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.276  0.862  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      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: 570"
## [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
## 157 104 105 107 134 80 79 65 96 85 91 121 124 184 162
## 2011 2012
## 152 142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 76 29 42 42 65 29 39 37 43 45 42 71 75 110 91
## 2011 2012

```

```
## 79 82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 27 35 34 47 26 34 33 40 41 33 61 69 94 78
## 2011 2012
## 65 68
## [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 = 1.2, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 2.78207445369764"
## [1] "Male first author team size 2018 geometric mean: 2.804844493199952"

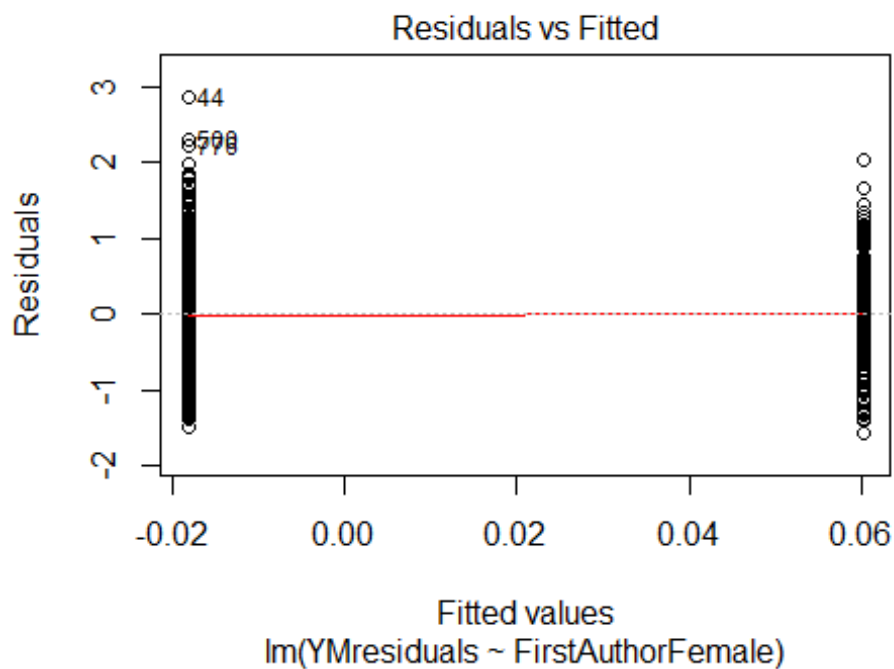
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.53917695148275"
## [1] "Male last author team size 2018 geometric mean: 2.85961289351258"

## 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 = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

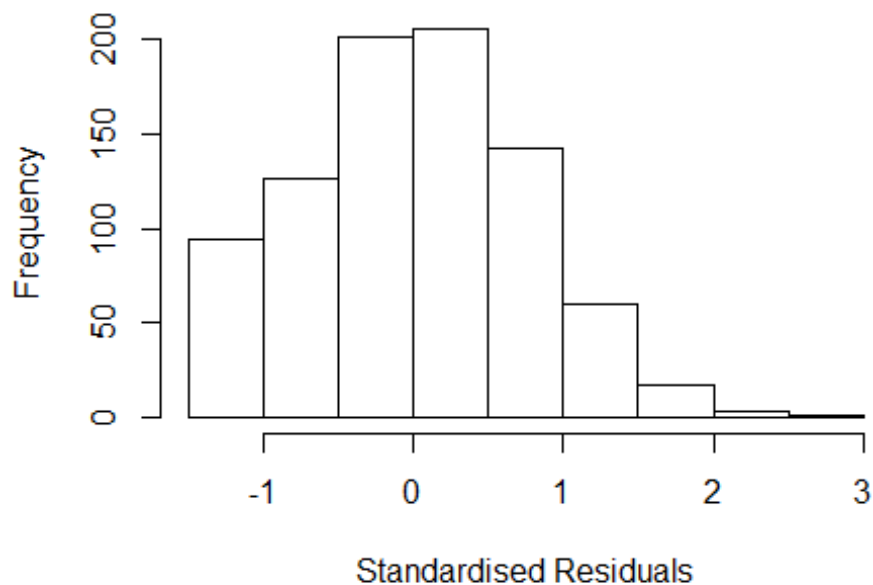
## Warning in lmrob.S(x, y, control = control, mf = mf): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps
```



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

```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.212  1      1.101
## LastAuthorFemale  1.214  1      1.102
## Year              1.141 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
## 44 0011083273 4.117 1996    1606      2      2.88
##
```

```
## Call:
```

```
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
```

```
## Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -1.47242 -0.52204  0.00759  0.52758  2.88014
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2369    0.0949   13.04  <2e-16 ***
## FirstAuthorFemale1  0.0968    0.0660    1.47    0.14
## LastAuthorFemale1 -0.0645    0.0752   -0.86    0.39
## Year1997          -0.0416    0.1695   -0.25    0.81
## Year1998          -0.2099    0.2060   -1.02    0.31
```

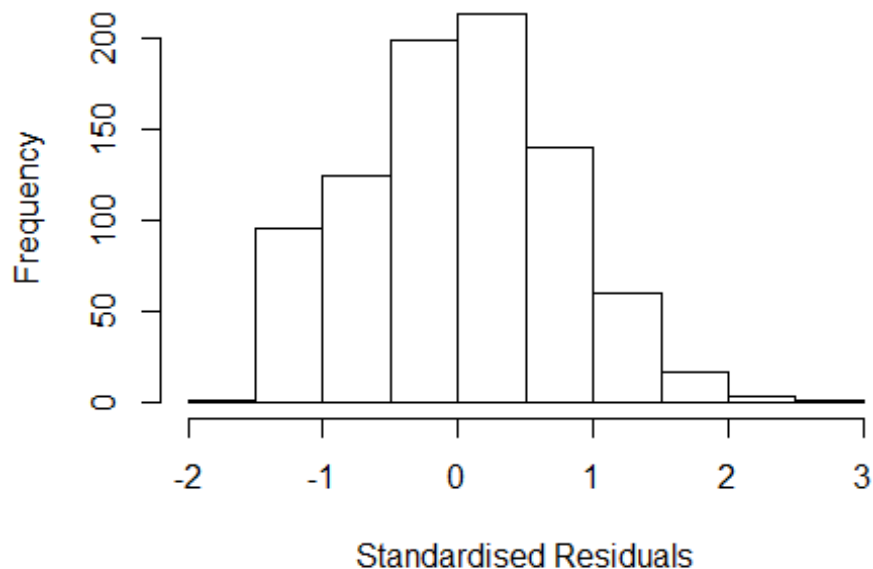


```

## Year1999      -0.2568      0.2195     -1.17      0.24
## Year2000      0.0013      0.1373      0.01      0.99
## Year2001     -0.0367      0.1803     -0.20      0.84
## Year2002     -0.0389      0.1619     -0.24      0.81
## Year2003     -0.0115      0.1630     -0.07      0.94
## Year2004      0.1788      0.1534      1.17      0.24
## Year2005      0.2033      0.1601      1.27      0.20
## Year2006     -0.1797      0.1651     -1.09      0.28
## Year2007      0.0668      0.1555      0.43      0.67
## Year2008      0.0104      0.1233      0.08      0.93
## Year2009      0.1376      0.1145      1.20      0.23
## Year2010     -0.1522      0.1361     -1.12      0.26
## Year2011     -0.0508      0.1321     -0.38      0.70
## Year2012     -0.0904      0.1298     -0.70      0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.777
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.00655
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 781 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.140  0.865  0.951  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      1.17e-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 1.071 1      1.035
## Year              1.071 16      1.002

```

Residuals from first author



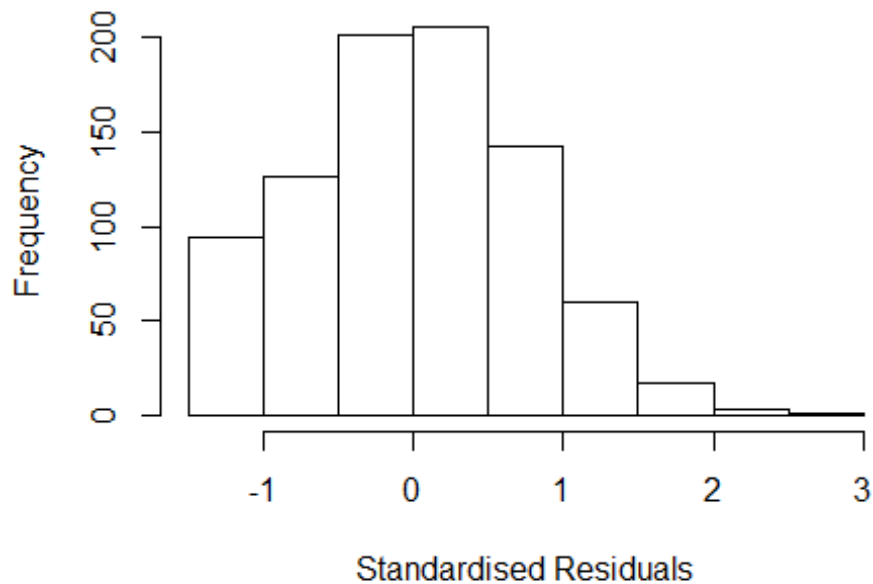
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 44 0011083273 4.117 1996      1606      2      2.88
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5096 -0.5185  0.0142  0.5216  2.8777
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.23925    0.09506   13.04  <2e-16 ***
## FirstAuthorFemale1 0.07569    0.06221    1.22    0.22
## Year1997        -0.04948    0.16817   -0.29    0.77
## Year1998        -0.21559    0.20624   -1.05    0.30
## Year1999        -0.27005    0.21719   -1.24    0.21
## Year2000        -0.01408    0.13665   -0.10    0.92
## Year2001        -0.04586    0.17883   -0.26    0.80
## Year2002        -0.04519    0.16202   -0.28    0.78
## Year2003        -0.02025    0.16339   -0.12    0.90
## Year2004         0.16851    0.15173    1.11    0.27
## Year2005         0.19462    0.16072    1.21    0.23
## Year2006        -0.18924    0.16444   -1.15    0.25
```

```

## Year2007          0.06074    0.15540    0.39    0.70
## Year2008         -0.00116    0.12227   -0.01    0.99
## Year2009          0.12775    0.11383    1.12    0.26
## Year2010         -0.16087    0.13603   -1.18    0.24
## Year2011         -0.06151    0.13167   -0.47    0.64
## Year2012         -0.09828    0.13004   -0.76    0.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.779
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.00679
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 783 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.143  0.865  0.953  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      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 1.066 1          1.033
## Year              1.066 16          1.002

```

Residuals from last author



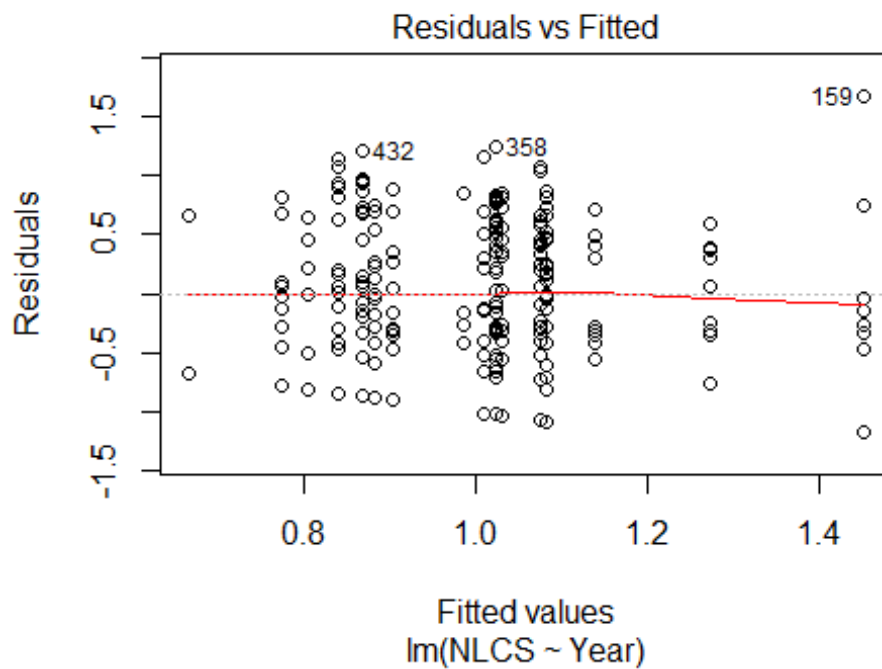
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 44 0011083273 4.117 1996      1606      2      2.88
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.45005 -0.51616  0.00889  0.52582  2.85989
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.25711    0.09487   13.25  <2e-16 ***
## LastAuthorFemale1 -0.02525    0.07006   -0.36    0.72
## Year1997        -0.06717    0.16775   -0.40    0.69
## Year1998        -0.22356    0.20725   -1.08    0.28
## Year1999        -0.27022    0.21776   -1.24    0.21
## Year2000        -0.00806    0.13763   -0.06    0.95
## Year2001        -0.04509    0.18328   -0.25    0.81
## Year2002        -0.03952    0.16345   -0.24    0.81
## Year2003        -0.01188    0.16352   -0.07    0.94
## Year2004         0.16727    0.15460    1.08    0.28
## Year2005         0.19294    0.16080    1.20    0.23
## Year2006        -0.18322    0.16492   -1.11    0.27
```

```

## Year2007      0.06653    0.15718    0.42    0.67
## Year2008      0.01492    0.12408    0.12    0.90
## Year2009      0.14056    0.11525    1.22    0.22
## Year2010     -0.15992    0.13603   -1.18    0.24
## Year2011     -0.05394    0.13203   -0.41    0.68
## Year2012     -0.08769    0.13065   -0.67    0.50
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.776
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.00523
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 783 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.145  0.862  0.951  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      1.17e-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: 852"
## [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
##   12   13   14   16   14   20   25   17   17   17   40   34   24   24   40
## 2011 2012
##   40   29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    2   11    9    6   10   12    9    8   11   24   25   17   15   29
## 2011 2012

```

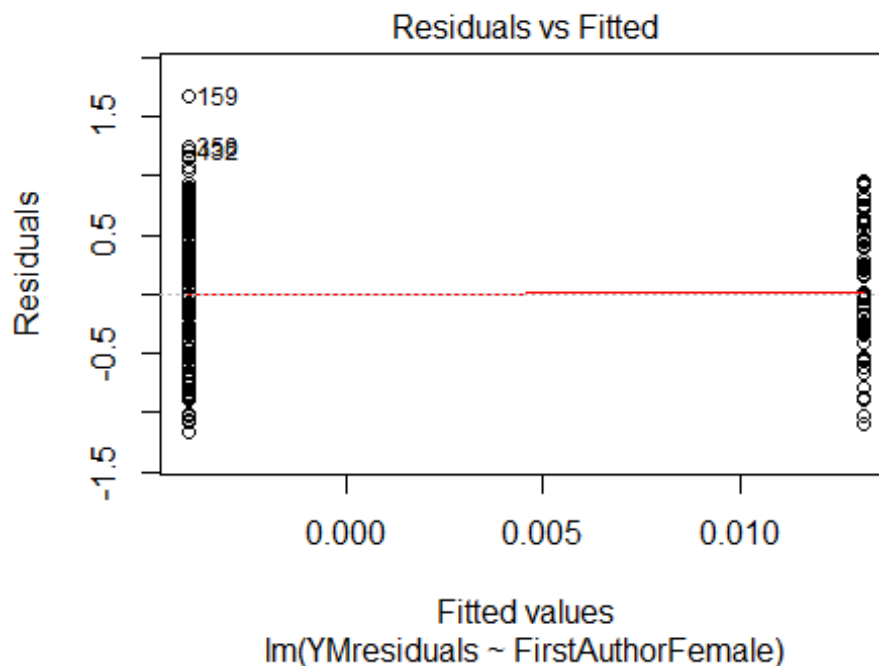
```
## 30 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 1 8 8 5 9 11 9 6 11 24 23 14 14 28
## 2011 2012
## 28 18
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 9.3, df = 16, p-value = 0.9
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.16, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 2.39492974229681"
## [1] "Male first author team size 2018 geometric mean: 2.61893333972043"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

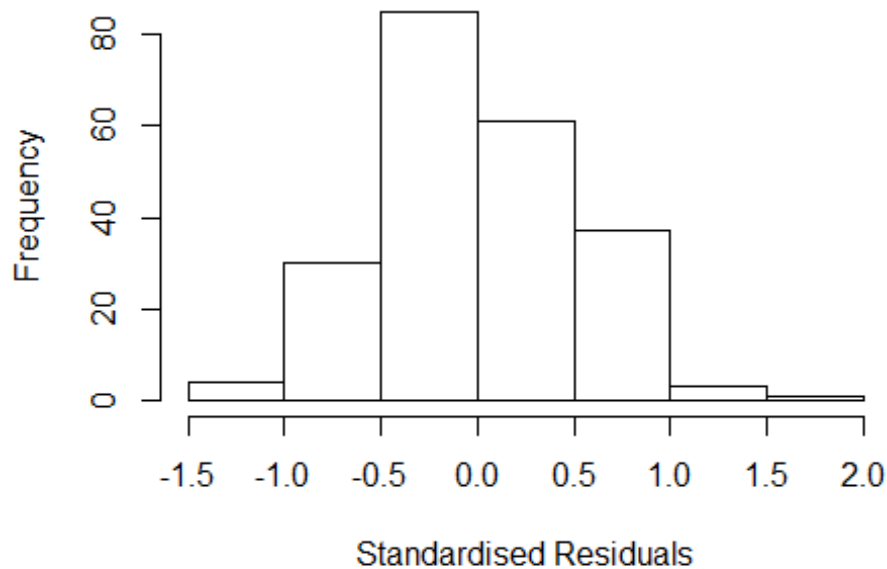
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 150, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.2837538219638"
## [1] "Male last author team size 2018 geometric mean: 2.65583388630574"

## 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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.274 1      1.508
## LastAuthorFemale 2.010 1      1.418
## UniqueAuthors    9.630 4      1.327
## Year             13.761 16     1.085
```

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
## -1.4388 -0.3440 -0.0633 0.3701 1.5091
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8402 0.2998 2.80 0.00557 **
## FirstAuthorFemale1 0.0616 0.0927 0.66 0.50684
## LastAuthorFemale1 -0.1509 0.1092 -1.38 0.16853
## UniqueAuthors2 0.1652 0.0925 1.79 0.07562 .
## UniqueAuthors3 0.6193 0.1287 4.81 3.0e-06 ***
## UniqueAuthors4 0.4365 0.1893 2.31 0.02215 *
## UniqueAuthors5 0.8619 0.1085 7.95 1.4e-13 ***
## Year1997 -1.0054 0.2850 -3.53 0.00052 ***
## Year1998 0.2348 0.3568 0.66 0.51125
## Year1999 -0.4069 0.3082 -1.32 0.18829
```

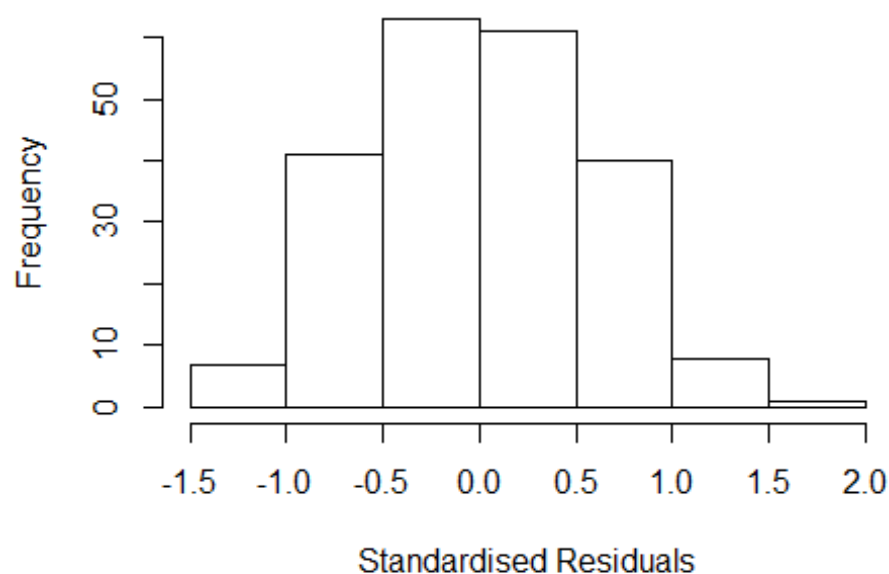


```

## Year2000          -0.1181      0.3659   -0.32  0.74720
## Year2001           0.1131      0.3562    0.32  0.75118
## Year2002           0.0189      0.3214    0.06  0.95313
## Year2003           0.0801      0.3316    0.24  0.80928
## Year2004          -0.0598      0.3716   -0.16  0.87240
## Year2005          -0.1913      0.3284   -0.58  0.56100
## Year2006          -0.0207      0.3121   -0.07  0.94712
## Year2007           0.0672      0.3212    0.21  0.83438
## Year2008          -0.1894      0.3217   -0.59  0.55672
## Year2009          -0.1009      0.3327   -0.30  0.76188
## Year2010          -0.1959      0.3101   -0.63  0.52832
## Year2011          -0.4002      0.3128   -1.28  0.20224
## Year2012          -0.4595      0.3214   -1.43  0.15439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.307, Adjusted R-squared:  0.23
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 206 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.412  0.878  0.955  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      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 2.062 1      1.436
## LastAuthorFemale  1.686 1      1.298
## Year              2.178 16      1.025

```

Residuals from first and last author



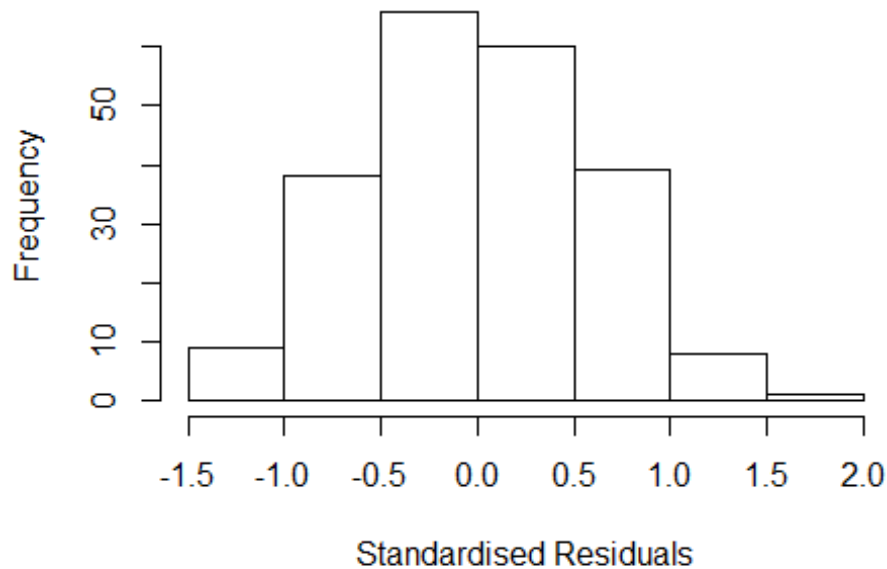
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10e+00 -3.96e-01 -1.67e-15 4.47e-01 1.94e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9586 0.2569 3.73 0.00025 ***
## FirstAuthorFemale1 0.1118 0.1010 1.11 0.26935
## LastAuthorFemale1 -0.2090 0.1421 -1.47 0.14283
## Year1997 -0.9586 0.2569 -3.73 0.00025 ***
## Year1998 0.2753 0.3082 0.89 0.37283
## Year1999 -0.2803 0.2902 -0.97 0.33527
## Year2000 -0.0299 0.3449 -0.09 0.93101
## Year2001 0.1763 0.3332 0.53 0.59741
## Year2002 -0.0163 0.3027 -0.05 0.95699
## Year2003 0.1617 0.2976 0.54 0.58753
## Year2004 0.2158 0.3788 0.57 0.56954
## Year2005 0.0456 0.3208 0.14 0.88717
```

```

## Year2006          0.1464      0.2865      0.51  0.60994
## Year2007          0.1378      0.2990      0.46  0.64530
## Year2008         -0.1269      0.2913     -0.44  0.66355
## Year2009          0.1908      0.3077      0.62  0.53603
## Year2010          0.0466      0.2860      0.16  0.87062
## Year2011         -0.1145      0.2987     -0.38  0.70198
## Year2012         -0.2603      0.3110     -0.84  0.40351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.675
## Multiple R-squared:  0.0781, Adjusted R-squared:  -0.00402
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 204 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.387  0.894  0.960  0.928  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 1.592 1          1.262
## Year              1.592 16          1.015

```

Residuals from first author



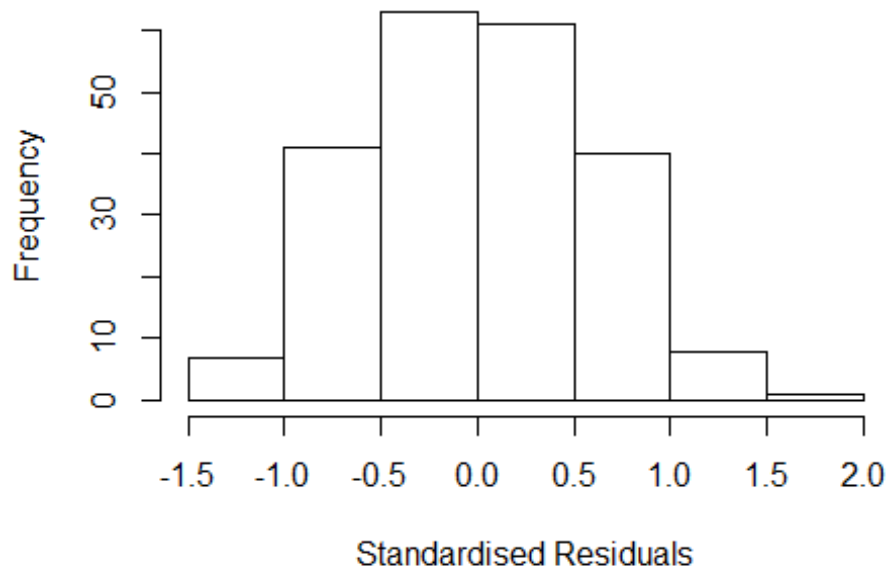
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1146 -0.3815 -0.0376 0.4647 1.9409
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9592 0.2567 3.74 0.00024 ***
## FirstAuthorFemale1 0.0311 0.0964 0.32 0.74699
## Year1997 -0.9592 0.2567 -3.74 0.00024 ***
## Year1998 0.2964 0.3082 0.96 0.33732
## Year1999 -0.2943 0.2944 -1.00 0.31875
## Year2000 -0.0439 0.3471 -0.13 0.89944
## Year2001 0.1782 0.3223 0.55 0.58096
## Year2002 -0.0305 0.3018 -0.10 0.91957
## Year2003 0.1641 0.3023 0.54 0.58787
## Year2004 0.2189 0.3806 0.58 0.56580
## Year2005 0.0301 0.3266 0.09 0.92655
## Year2006 0.1243 0.2870 0.43 0.66542
```

```

## Year2007          0.0923      0.2936      0.31  0.75351
## Year2008         -0.1280      0.2918     -0.44  0.66142
## Year2009          0.1447      0.3092      0.47  0.64026
## Year2010          0.0410      0.2861      0.14  0.88611
## Year2011         -0.1300      0.2989     -0.44  0.66397
## Year2012         -0.2515      0.3074     -0.82  0.41428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.681
## Multiple R-squared:  0.0672, Adjusted R-squared:  -0.0109
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 205 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.397  0.894  0.953  0.929  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 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.389 1          1.178
## Year              1.389 16          1.010

```

Residuals from last author



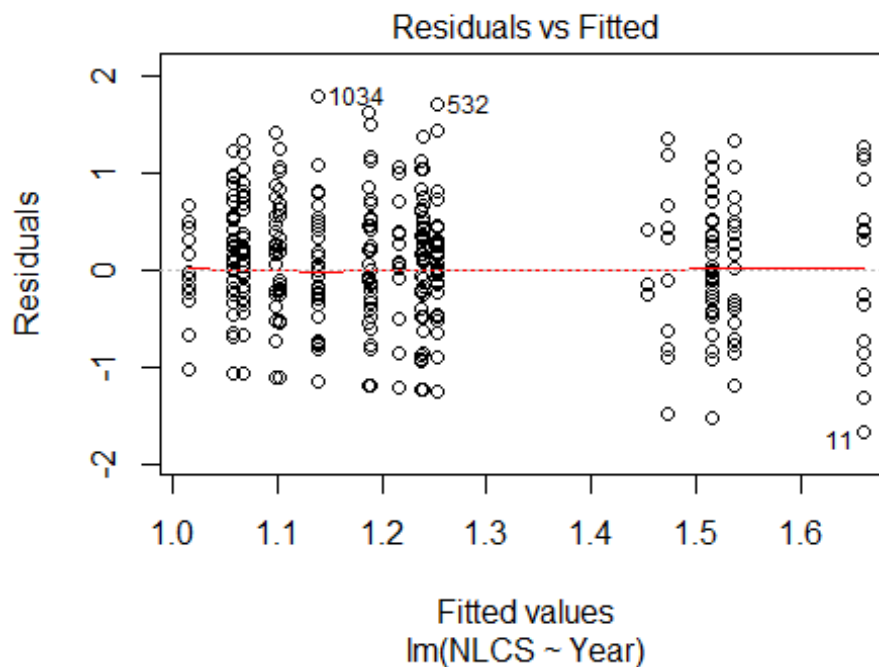
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13276 -0.38721 -0.00878 0.45413 1.94693
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95821 0.25681 3.73 0.00025 ***
## LastAuthorFemale1 -0.14601 0.12599 -1.16 0.24785
## Year1997 -0.95821 0.25681 -3.73 0.00025 ***
## Year1998 0.30585 0.30654 1.00 0.31959
## Year1999 -0.27457 0.28872 -0.95 0.34273
## Year2000 0.00495 0.35169 0.01 0.98879
## Year2001 0.20029 0.32417 0.62 0.53737
## Year2002 -0.01082 0.30344 -0.04 0.97158
## Year2003 0.19311 0.30050 0.64 0.52118
## Year2004 0.21386 0.37777 0.57 0.57195
## Year2005 0.06782 0.31846 0.21 0.83157
## Year2006 0.17455 0.28624 0.61 0.54267
```

```

## Year2007      0.14205      0.29852      0.48  0.63468
## Year2008     -0.10343      0.29078     -0.36  0.72245
## Year2009      0.20370      0.30655      0.66  0.50713
## Year2010      0.07553      0.28539      0.26  0.79154
## Year2011     -0.09334      0.30003     -0.31  0.75604
## Year2012     -0.23865      0.30789     -0.78  0.43917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.671
## Multiple R-squared:  0.0742, Adjusted R-squared:  -0.00334
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 209 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.380  0.890  0.959  0.929  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] "Sample size for the above analysis: 221"
## [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
##   46   27   34   39   41   46   37   49   44   45   57   60   68   83   75
## 2011 2012
##   65   73
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15    3   11   13   17   10   15   23   21   22   31   27   30   44   49
## 2011 2012

```

```
## 32 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 3 11 13 15 8 15 20 18 21 29 26 28 37 44
## 2011 2012
## 27 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 = 13, df = 16, p-value = 0.7
```

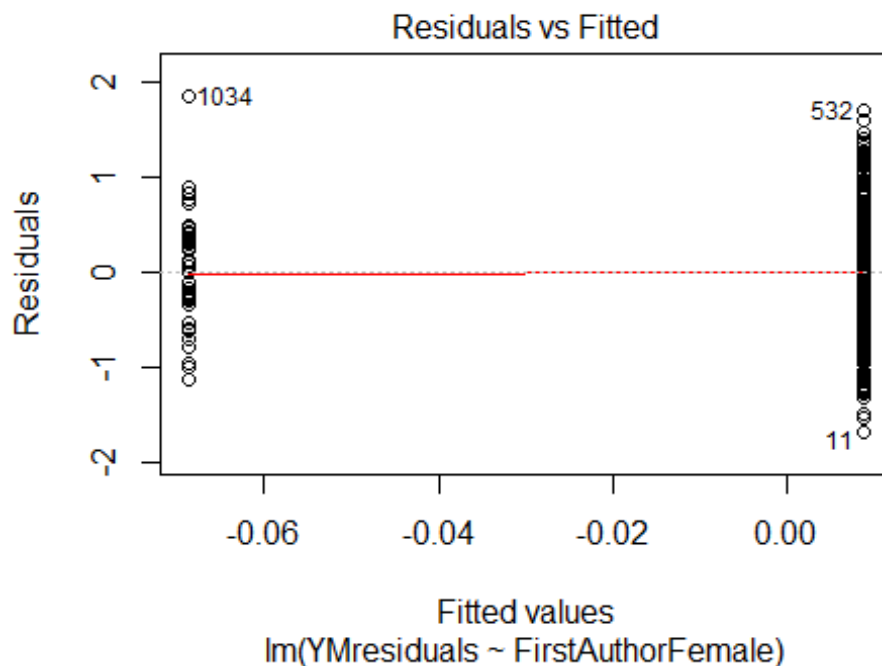


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.3
## [1] "Female first author team size 2018 geometric mean: 1.8612097182042"
## [1] "Male first author team size 2018 geometric mean: 2.57257186892704"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



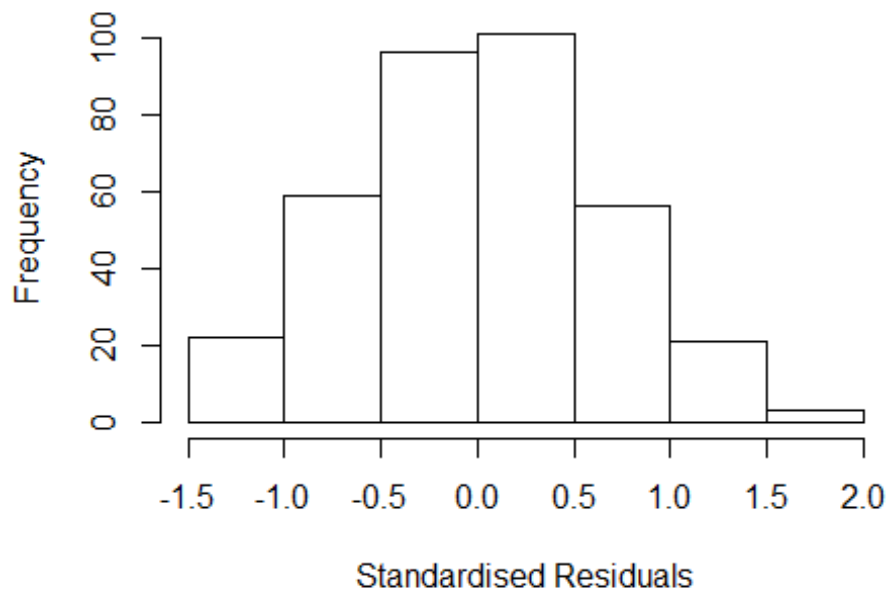
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 40, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1"
## [1] "Male last author team size 2018 geometric mean: 2.54698805190545"

## 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 = 0.2
## 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 1.612  1      1.270
## LastAuthorFemale  1.365  1      1.168
## UniqueAuthors    2.218  4      1.105
## Year              3.068 16      1.036
```

Residuals from first and last author and team size



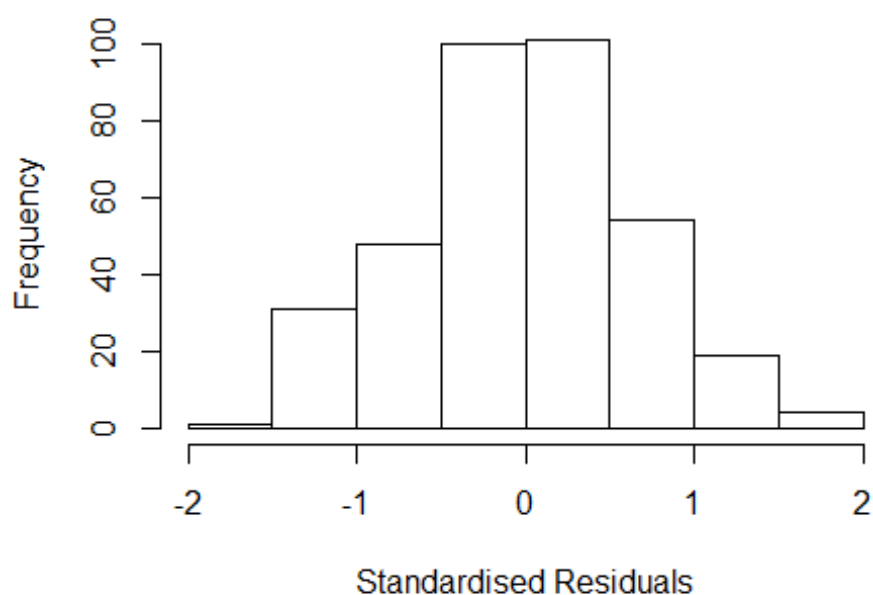
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4996 -0.4228 0.0129 0.4432 1.8719
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.49959 0.27906 5.37 1.4e-07 ***
## FirstAuthorFemale1 -0.10746 0.11642 -0.92 0.35665
## LastAuthorFemale1 -0.00642 0.12334 -0.05 0.95851
## UniqueAuthors2 0.32140 0.09315 3.45 0.00063 ***
## UniqueAuthors3 0.54347 0.10026 5.42 1.1e-07 ***
## UniqueAuthors4 0.38105 0.15878 2.40 0.01695 *
## UniqueAuthors5 0.29536 0.14227 2.08 0.03865 *
## Year1997 -0.15512 0.37137 -0.42 0.67644
## Year1998 -0.56450 0.35637 -1.58 0.11413
## Year1999 -0.59400 0.30792 -1.93 0.05457 .
```

```

## Year2000      -0.52691      0.33772      -1.56      0.11966
## Year2001      -0.36773      0.57079      -0.64      0.51986
## Year2002      -0.35871      0.32875      -1.09      0.27600
## Year2003      -0.60661      0.31111      -1.95      0.05203 .
## Year2004      -0.22142      0.30449      -0.73      0.46763
## Year2005      -0.47551      0.30861      -1.54      0.12431
## Year2006      -0.44431      0.30166      -1.47      0.14172
## Year2007      -0.53282      0.29902      -1.78      0.07567 .
## Year2008      -0.61585      0.29873      -2.06      0.04002 *
## Year2009      -0.70604      0.29002      -2.43      0.01544 *
## Year2010      -0.72463      0.28613      -2.53      0.01178 *
## Year2011      -0.62359      0.29580      -2.11      0.03576 *
## Year2012      -0.41346      0.29173      -1.42      0.15733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.685
## Multiple R-squared:  0.138, Adjusted R-squared:  0.081
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 325 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.436  0.882  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          2.79e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.549 1          1.245
## LastAuthorFemale  1.291 1          1.136
## Year              1.524 16          1.013

```

Residuals from first and last author



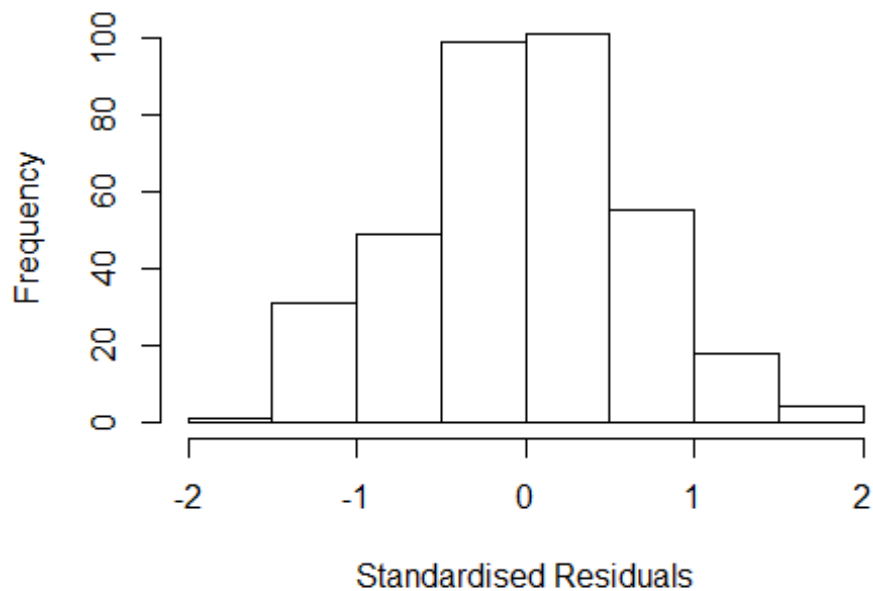
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.69539 -0.44620 -0.00226 0.44585 1.87899
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.695 0.342 4.96 1.1e-06 ***
## FirstAuthorFemale1 -0.084 0.115 -0.73 0.466
## LastAuthorFemale1 -0.065 0.127 -0.51 0.608
## Year1997 -0.245 0.385 -0.64 0.525
## Year1998 -0.512 0.421 -1.22 0.224
## Year1999 -0.656 0.366 -1.79 0.074 .
## Year2000 -0.561 0.398 -1.41 0.160
## Year2001 -0.264 0.515 -0.51 0.609
## Year2002 -0.453 0.388 -1.17 0.244
## Year2003 -0.542 0.386 -1.40 0.162
## Year2004 -0.239 0.371 -0.64 0.520
## Year2005 -0.605 0.372 -1.62 0.105
```

```

## Year2006          -0.466      0.363   -1.29    0.200
## Year2007          -0.552      0.364   -1.52    0.130
## Year2008          -0.493      0.362   -1.36    0.174
## Year2009          -0.712      0.356   -2.00    0.046 *
## Year2010          -0.650      0.356   -1.83    0.068 .
## Year2011          -0.566      0.362   -1.57    0.118
## Year2012          -0.248      0.352   -0.70    0.482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.695
## Multiple R-squared:  0.0647, Adjusted R-squared:  0.0151
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 333 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.445  0.873  0.959  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      2.79e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.288 1      1.135
## Year              1.288 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
## -1.69558 -0.44110 -0.00152 0.45136 1.90633
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.696 0.344 4.93 1.3e-06 ***
## FirstAuthorFemale1 -0.111 0.107 -1.03 0.303
## Year1997 -0.245 0.386 -0.63 0.526
## Year1998 -0.530 0.418 -1.27 0.205
## Year1999 -0.657 0.367 -1.79 0.074 .
## Year2000 -0.559 0.400 -1.40 0.163
## Year2001 -0.259 0.518 -0.50 0.617
## Year2002 -0.453 0.390 -1.16 0.246
## Year2003 -0.551 0.386 -1.43 0.155
## Year2004 -0.237 0.372 -0.64 0.524
## Year2005 -0.605 0.374 -1.62 0.107
## Year2006 -0.468 0.364 -1.28 0.200
```

```

## Year2007          -0.555      0.366   -1.52    0.130
## Year2008          -0.503      0.362   -1.39    0.166
## Year2009          -0.718      0.358   -2.01    0.045 *
## Year2010          -0.655      0.357   -1.84    0.067 .
## Year2011          -0.567      0.363   -1.56    0.120
## Year2012          -0.246      0.354   -0.69    0.488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.694
## Multiple R-squared:  0.0643, Adjusted R-squared:  0.0175
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 331 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.430  0.871  0.959  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      2.79e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.099 1      1.048
## Year      1.099 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
## -1.68190 -0.43761 -0.00146  0.43824  1.80831

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.682      0.345   4.88 1.7e-06 ***
## LastAuthorFemale1 -0.107      0.119  -0.89  0.373
## Year1997          -0.231      0.387  -0.60  0.551
## Year1998          -0.492      0.423  -1.16  0.246
## Year1999          -0.648      0.370  -1.75  0.081 .
## Year2000          -0.555      0.403  -1.38  0.169
## Year2001          -0.264      0.514  -0.51  0.608
## Year2002          -0.439      0.391  -1.12  0.262
## Year2003          -0.525      0.387  -1.36  0.176
## Year2004          -0.231      0.375  -0.62  0.539
## Year2005          -0.591      0.375  -1.58  0.116
## Year2006          -0.455      0.366  -1.24  0.215
## Year2007          -0.547      0.369  -1.48  0.140
## Year2008          -0.472      0.363  -1.30  0.194
## Year2009          -0.708      0.362  -1.96  0.051 .
## Year2010          -0.645      0.361  -1.79  0.075 .
## Year2011          -0.566      0.368  -1.54  0.125
## Year2012          -0.254      0.359  -0.71  0.479
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.693
## Multiple R-squared:  0.0633, Adjusted R-squared:  0.0164
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 327 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.475  0.872  0.959  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           2.79e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 358"
## [1] ""

```



```

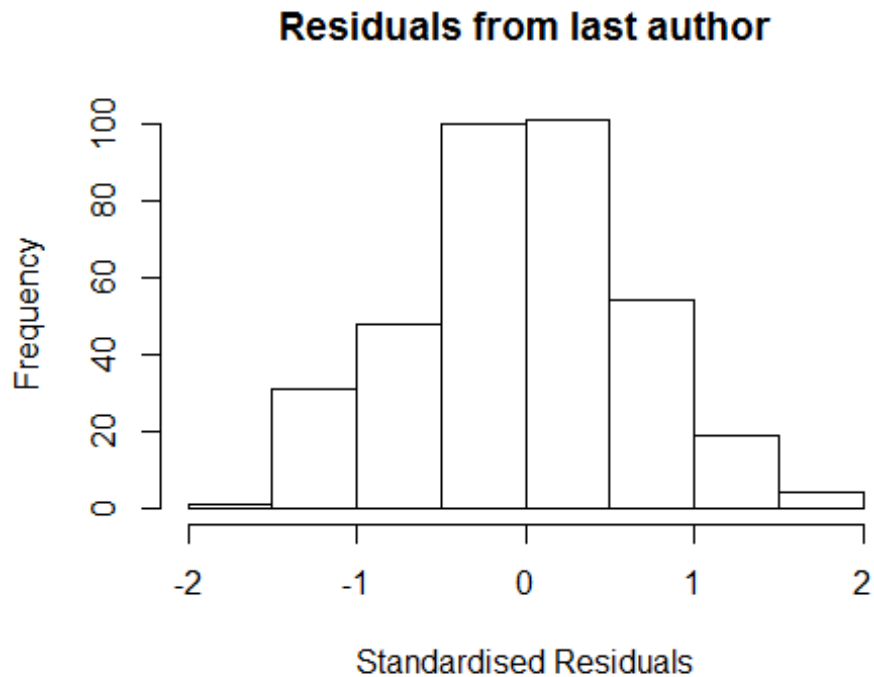
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2203"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    3    7   10   15   10   14    6   12    8   20   14   25   36   31
## 2011 2012
##   35   47
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    1    2    6   10    4   12    1    8    4   12    7   12   16   14
## 2011 2012
##   19   21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    1    2    6    9    4   12    1    6    4   11    6   11   13   11
## 2011 2012
##   15   15
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.94755852705582"
## [1] "Male first author team size 2018 geometric mean: 2.22888344742581"

## 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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.00899287331754"
## [1] "Male last author team size 2018 geometric mean: 2.32754366950671"

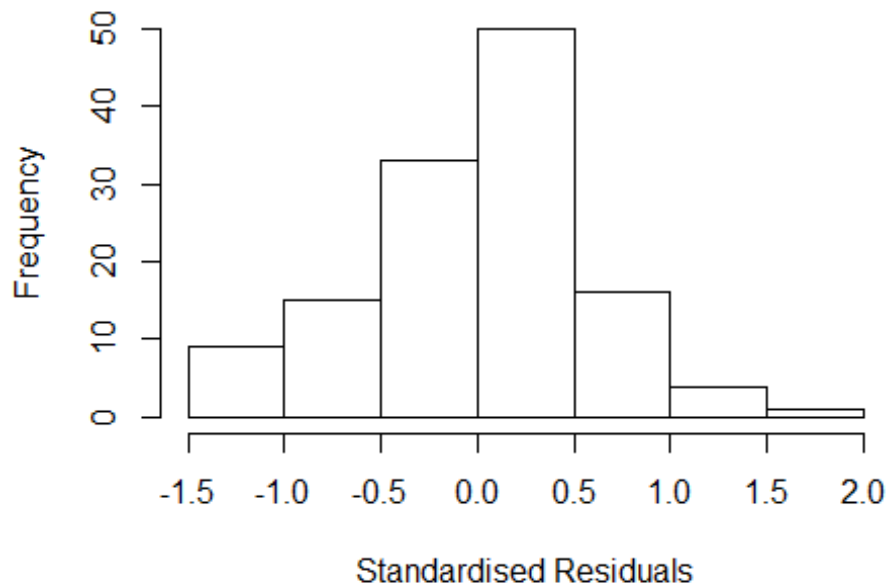
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.4
## 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 2.170e+00  1          1.473
## LastAuthorFemale  1.562e+00  1          1.250
## UniqueAuthors    4.416e+26  4         2141.088
## Year              7.615e+26 16          6.919
```

Residuals from first and last author and team size



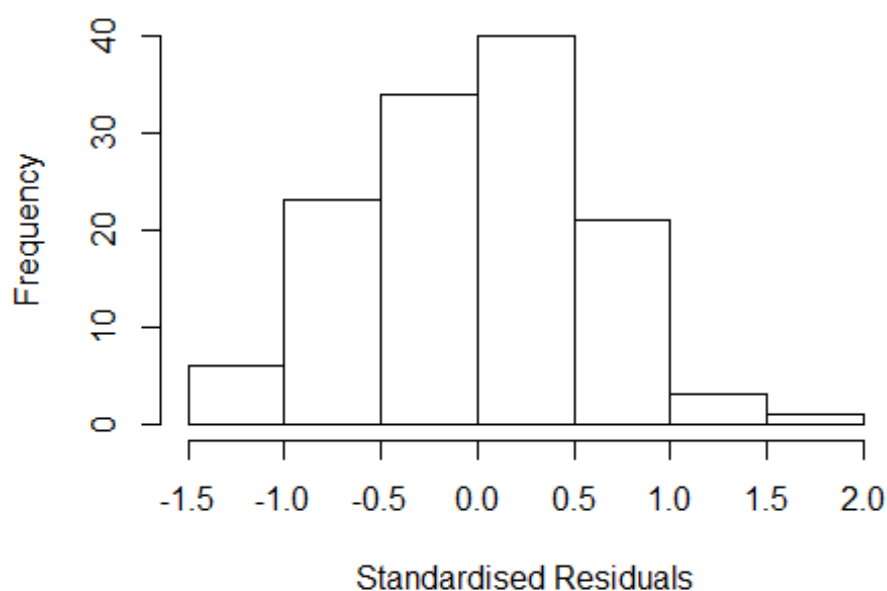
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4519 -0.3885 0.0523 0.3383 1.7936
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.40e+00 1.02e-08 1.38e+08 < 2e-16 ***
## FirstAuthorFemale1 2.34e-02 1.37e-01 1.70e-01 0.86496
## LastAuthorFemale1 -1.15e-01 2.20e-01 -5.20e-01 0.60130
## UniqueAuthors2 2.91e-01 1.82e-01 1.60e+00 0.11171
## UniqueAuthors3 5.55e-01 1.71e-01 3.25e+00 0.00156 **
## UniqueAuthors4 4.33e-01 1.59e-01 2.73e+00 0.00744 **
## UniqueAuthors5 1.37e-02 2.56e-01 5.00e-02 0.95742
## Year1997 5.35e-01 1.82e-01 2.94e+00 0.00400 **
## Year1998 8.60e-02 4.82e-01 1.80e-01 0.85879
## Year1999 -3.37e-01 1.51e-01 -2.24e+00 0.02753 *
```

```

## Year2000      -1.65e-01    2.49e-01 -6.60e-01    0.51055
## Year2001      -1.71e-01    1.72e-01 -1.00e+00    0.32087
## Year2002       1.36e-01    3.20e-01  4.20e-01    0.67210
## Year2003      -2.72e-01    2.56e-01 -1.06e+00    0.29057
## Year2004      -5.35e-02    1.88e-01 -2.80e-01    0.77625
## Year2005      -8.56e-03    4.89e-01 -2.00e-02    0.98605
## Year2006      -1.43e-01    2.28e-01 -6.30e-01    0.53183
## Year2007      -3.78e-01    2.38e-01 -1.59e+00    0.11514
## Year2008      -8.51e-01    2.20e-01 -3.87e+00    0.00019 ***
## Year2009      -5.04e-01    3.28e-01 -1.53e+00    0.12791
## Year2010      -3.71e-01    2.03e-01 -1.83e+00    0.07067 .
## Year2011      -2.07e-01    1.74e-01 -1.19e+00    0.23666
## Year2012      -8.01e-01    2.21e-01 -3.62e+00    0.00045 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.27,    Adjusted R-squared:  0.117
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.233  0.866  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          7.81e-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 2.273 1          1.508
## LastAuthorFemale  1.258 1          1.122
## Year              2.646 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
## -1.3837 -0.3031 0.0221 0.4394 1.7933
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.40e+00 8.08e-08 1.73e+07 <2e-16 ***
## FirstAuthorFemale1 6.57e-02 1.47e-01 4.50e-01 0.657
## LastAuthorFemale1 -1.48e-01 2.08e-01 -7.10e-01 0.478
## Year1997 8.26e-01 8.41e-08 9.82e+06 <2e-16 ***
## Year1998 8.60e-02 4.58e-01 1.90e-01 0.851
## Year1999 -2.88e-01 1.37e-01 -2.10e+00 0.038 *
## Year2000 -1.73e-02 2.85e-01 -6.00e-02 0.952
## Year2001 1.79e-01 1.01e-01 1.77e+00 0.080 .
## Year2002 1.43e-01 2.94e-01 4.80e-01 0.629
## Year2003 -2.58e-01 8.21e-08 -3.14e+06 <2e-16 ***
## Year2004 1.98e-01 2.10e-01 9.40e-01 0.348
## Year2005 1.21e-01 5.05e-01 2.40e-01 0.812
```

```

## Year2006          1.02e-01  1.63e-01  6.20e-01  0.534
## Year2007          -1.42e-01  2.23e-01 -6.40e-01  0.526
## Year2008          -5.70e-01  2.22e-01 -2.57e+00  0.012 *
## Year2009          -1.86e-01  2.48e-01 -7.50e-01  0.455
## Year2010          -1.11e-02  1.75e-01 -6.00e-02  0.950
## Year2011           1.04e-01  1.03e-01  1.01e+00  0.315
## Year2012          -5.09e-01  2.08e-01 -2.45e+00  0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.173, Adjusted R-squared:  0.0365
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.345  0.854  0.944  0.905  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      7.81e-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 2.157 1      1.469
## Year              2.157 16      1.024
##
## [1] "List of 0 outliers 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

```

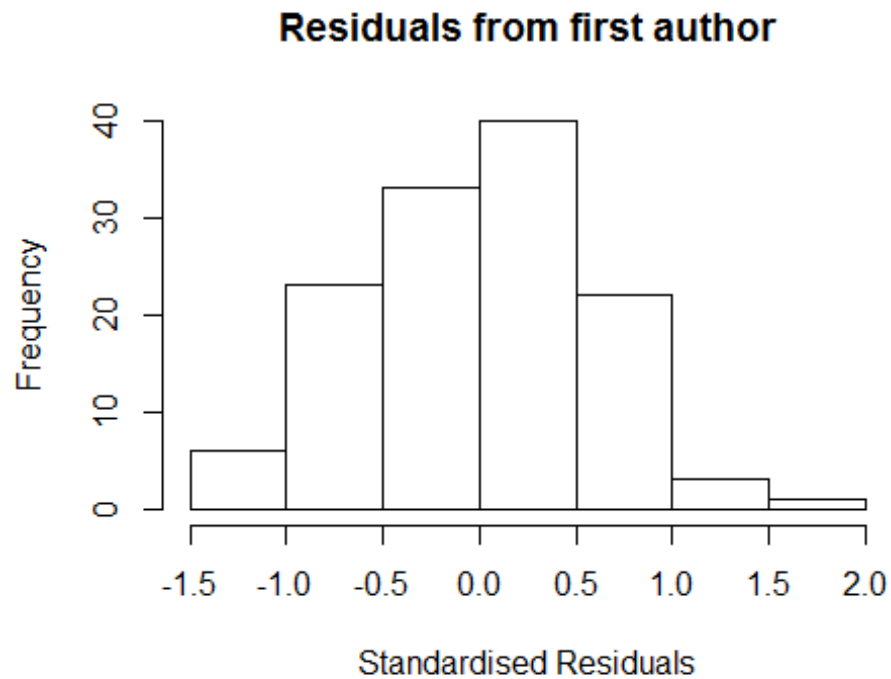
```

## -1.3799 -0.3113 0.0215 0.4419 1.8055
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.40e+00   9.41e-08  1.49e+07 <2e-16 ***
## FirstAuthorFemale1 2.83e-02   1.52e-01  1.90e-01  0.853
## Year1997        8.26e-01   1.00e-07  8.23e+06 <2e-16 ***
## Year1998        8.60e-02   4.53e-01  1.90e-01  0.850
## Year1999       -2.89e-01   1.37e-01 -2.10e+00  0.038 *
## Year2000       -2.11e-02   2.82e-01 -7.00e-02  0.940
## Year2001        1.89e-01   1.01e-01  1.87e+00  0.064 .
## Year2002        1.36e-01   2.96e-01  4.60e-01  0.648
## Year2003       -2.58e-01   9.56e-08 -2.70e+06 <2e-16 ***
## Year2004        2.04e-01   2.11e-01  9.70e-01  0.336
## Year2005        1.23e-01   4.94e-01  2.50e-01  0.803
## Year2006        9.34e-02   1.59e-01  5.90e-01  0.559
## Year2007       -1.40e-01   2.19e-01 -6.40e-01  0.523
## Year2008       -5.89e-01   2.25e-01 -2.62e+00  0.010 *
## Year2009       -2.04e-01   2.45e-01 -8.30e-01  0.406
## Year2010       -2.88e-02   1.76e-01 -1.60e-01  0.870
## Year2011        1.02e-01   1.03e-01  9.80e-01  0.327
## Year2012       -5.22e-01   2.02e-01 -2.59e+00  0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.167, Adjusted R-squared:  0.0388
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 113 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.367 0.847 0.943 0.908 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           7.81e-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"

```

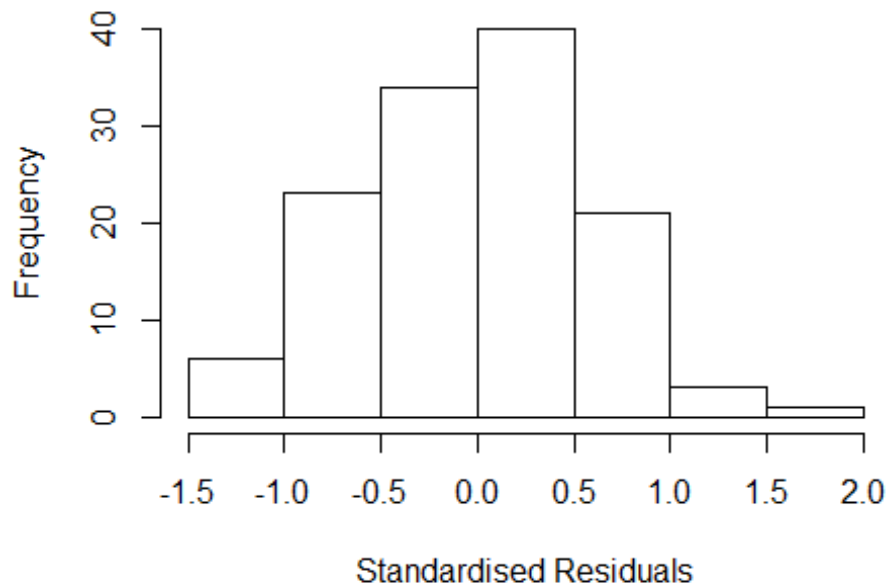
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## 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 16          NaN
```


Residuals from last author



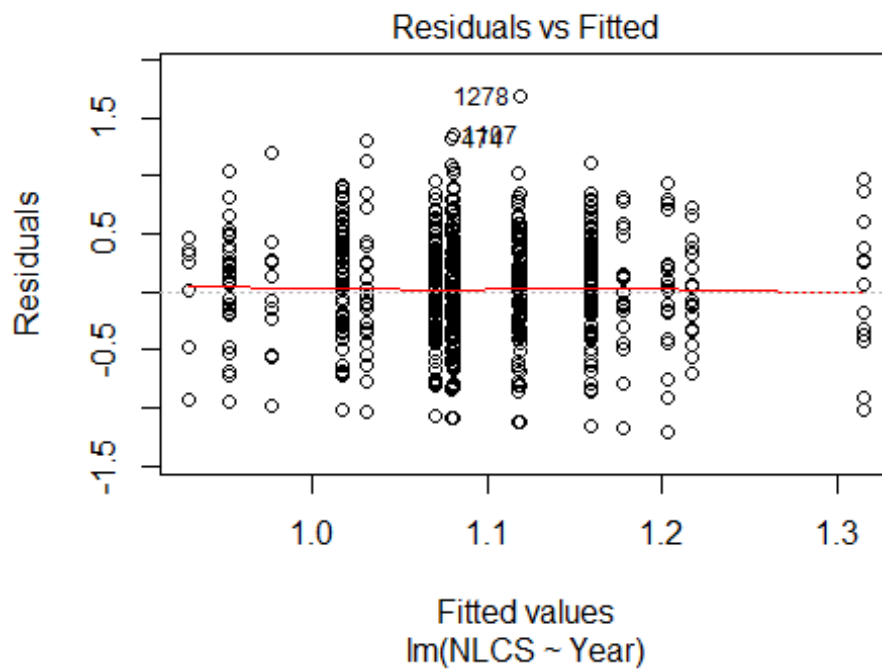
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.386 -0.315 0.028 0.431 1.787
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.401000 0.000000 Inf <2e-16 ***
## LastAuthorFemale1 -0.120814 0.206985 -0.58 0.5606
## Year1997 0.826000 0.000000 Inf <2e-16 ***
## Year1998 0.086000 0.461126 0.19 0.8524
## Year1999 -0.287012 0.136903 -2.10 0.0383 *
## Year2000 -0.015000 0.287114 -0.05 0.9584
## Year2001 0.195907 0.090916 2.15 0.0334 *
## Year2002 0.144820 0.295184 0.49 0.6247
## Year2003 -0.258000 0.000000 -Inf <2e-16 ***
## Year2004 0.211725 0.207572 1.02 0.3100
## Year2005 0.119111 0.512510 0.23 0.8167
## Year2006 0.129156 0.143429 0.90 0.3698
```

```

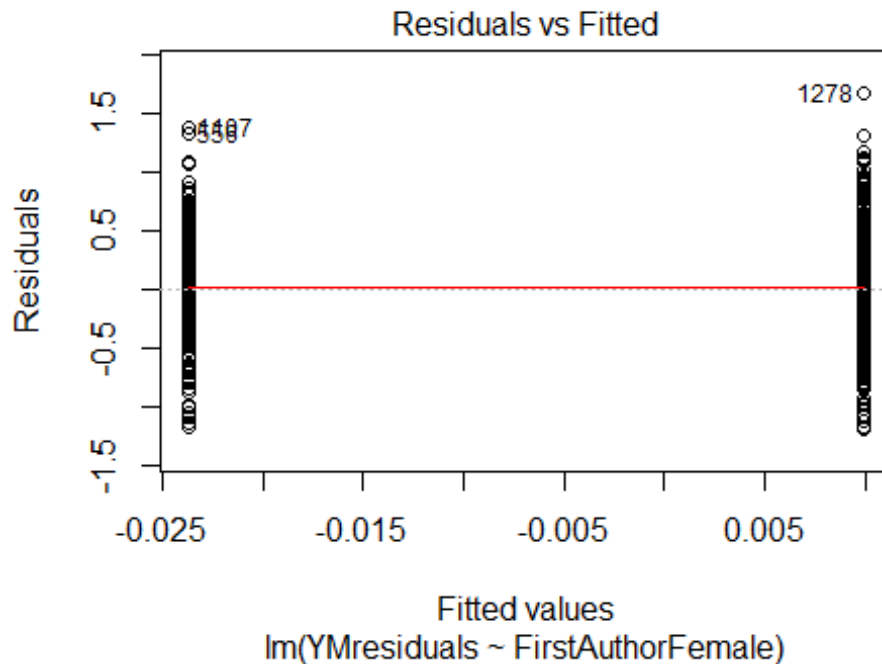
## Year2007          -0.126701    0.218116    -0.58    0.5625
## Year2008          -0.550624    0.204276    -2.70    0.0081 **
## Year2009          -0.181009    0.254454    -0.71    0.4784
## Year2010           0.000852    0.179971     0.00    0.9962
## Year2011           0.116505    0.094897     1.23    0.2222
## Year2012          -0.503245    0.214224    -2.35    0.0206 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.173, Adjusted R-squared:  0.0455
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.331  0.850  0.943  0.902  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      7.81e-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: 128"
## [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
##   46   41   69   56   44   64   40   53   57   78   76  114  145  136  208
## 2011 2012
##  169  188
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19    7   19   18   14   21   12   24   25   38   40   51   78   64  116
## 2011 2012

```

```
## 90 104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 5 13 16 13 18 10 17 15 28 35 38 48 48 103
## 2011 2012
## 82 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 = 12, df = 16, p-value = 0.8
```

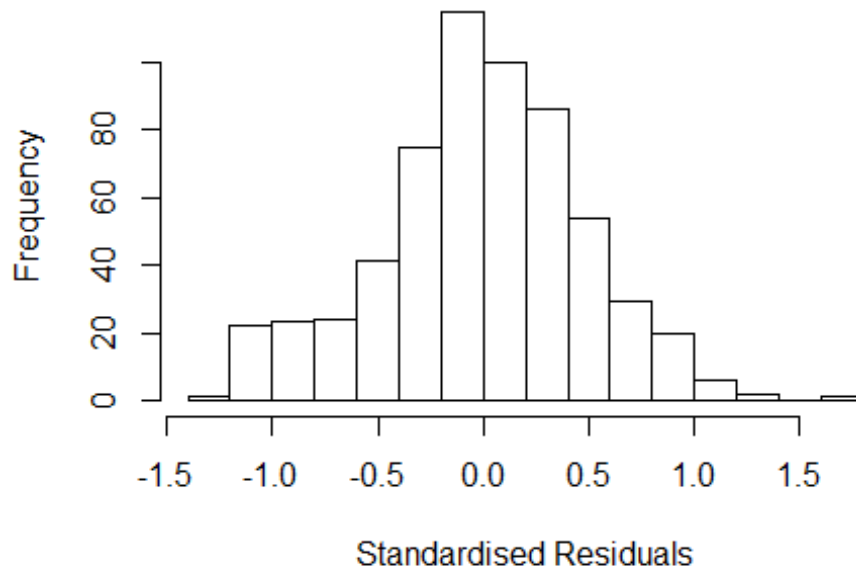


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 5.05750143299245"
## [1] "Male first author team size 2018 geometric mean: 4.51625482629232"
##
## 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: 4.41677554972109"
## [1] "Male last author team size 2018 geometric mean: 4.73702391794265"
##
## 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 1.313 1          1.146
## LastAuthorFemale  1.384 1          1.177
## UniqueAuthors    1.523 4          1.054
## Year              2.067 16         1.023
```

Residuals from first and last author and team size



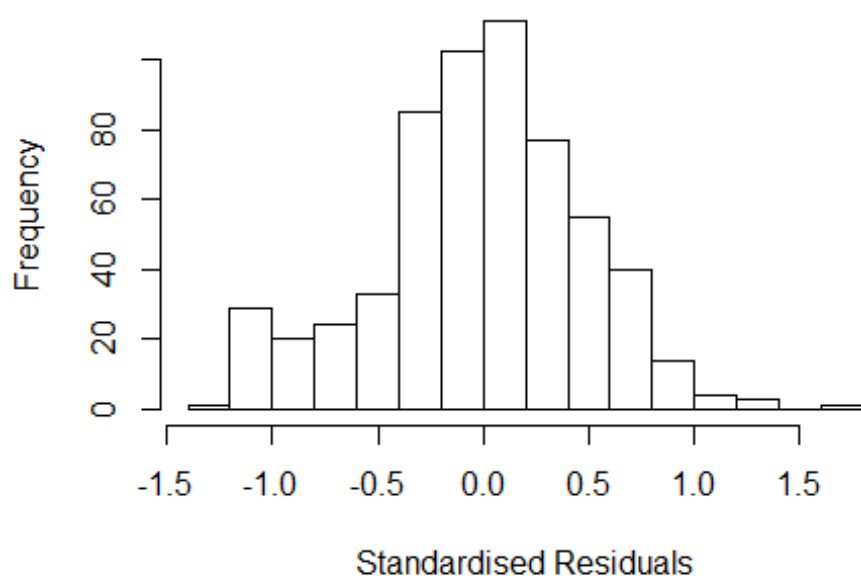
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31437 -0.27273 -0.00423  0.30440  1.67767
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13336    0.19882    5.70 1.9e-08 ***
## FirstAuthorFemale1 -0.02029    0.05032   -0.40 0.68693
## LastAuthorFemale1 -0.00761    0.05926   -0.13 0.89790
## UniqueAuthors2     0.15317    0.09384    1.63 0.10317
## UniqueAuthors3     0.18098    0.08492    2.13 0.03350 *
## UniqueAuthors4     0.30129    0.08889    3.39 0.00075 ***
## UniqueAuthors5     0.20131    0.08383    2.40 0.01665 *
## Year1997         -0.34793    0.37952   -0.92 0.35965
## Year1998         -0.08801    0.23144   -0.38 0.70388
## Year1999         -0.13845    0.22248   -0.62 0.53399
```

```

## Year2000      -0.01622      0.29071      -0.06      0.95552
## Year2001      -0.24275      0.21985      -1.10      0.26998
## Year2002      -0.23074      0.25971      -0.89      0.37468
## Year2003      -0.19433      0.22575      -0.86      0.38971
## Year2004      -0.18952      0.23702      -0.80      0.42428
## Year2005      -0.26339      0.22321      -1.18      0.23848
## Year2006      -0.18040      0.21461      -0.84      0.40091
## Year2007      -0.13354      0.21058      -0.63      0.52624
## Year2008      -0.15614      0.20153      -0.77      0.43879
## Year2009      -0.19202      0.19909      -0.96      0.33521
## Year2010      -0.23098      0.19948      -1.16      0.24738
## Year2011      -0.29409      0.20553      -1.43      0.15301
## Year2012      -0.15621      0.19938      -0.78      0.43366
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.451
## Multiple R-squared:  0.0456, Adjusted R-squared:  0.00916
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 539 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.136  0.855  0.952   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      1.67e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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  1.313 1      1.146
## Year              1.343 16      1.009

```

Residuals from first and last author



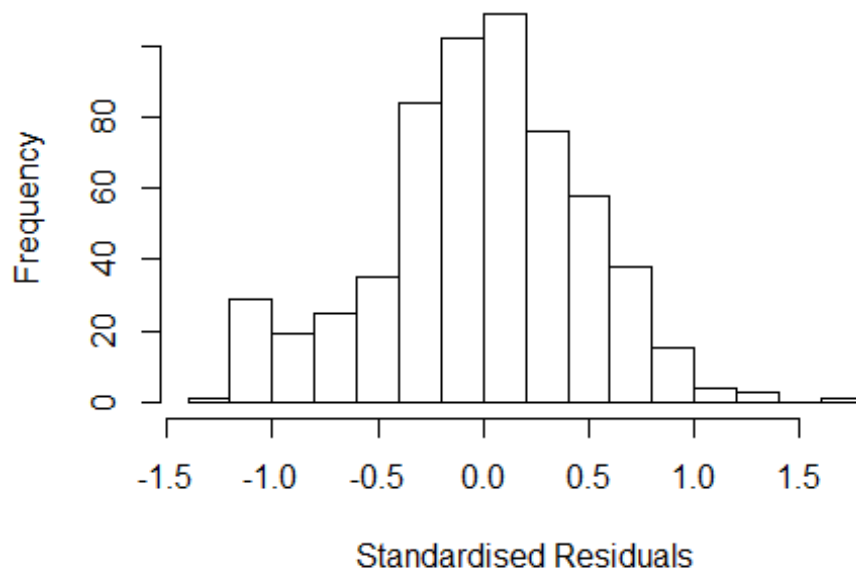
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2433 -0.2958 0.0126 0.3001 1.6742
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25183 0.17169 7.29 1e-12 ***
## FirstAuthorFemale1 -0.00858 0.05025 -0.17 0.86
## LastAuthorFemale1 -0.01900 0.05967 -0.32 0.75
## Year1997 -0.33818 0.33992 -0.99 0.32
## Year1998 -0.07052 0.21518 -0.33 0.74
## Year1999 -0.06821 0.19959 -0.34 0.73
## Year2000 0.06288 0.26502 0.24 0.81
## Year2001 -0.19295 0.20462 -0.94 0.35
## Year2002 -0.20457 0.24326 -0.84 0.40
## Year2003 -0.14363 0.21290 -0.67 0.50
## Year2004 -0.16521 0.22563 -0.73 0.46
## Year2005 -0.22095 0.19999 -1.10 0.27
```

```

## Year2006      -0.11240    0.19627   -0.57    0.57
## Year2007      -0.07621    0.19078   -0.40    0.69
## Year2008      -0.08204    0.17863   -0.46    0.65
## Year2009      -0.12606    0.17490   -0.72    0.47
## Year2010      -0.15411    0.17449   -0.88    0.38
## Year2011      -0.22586    0.18106   -1.25    0.21
## Year2012      -0.07804    0.17422   -0.45    0.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.462
## Multiple R-squared:  0.0194, Adjusted R-squared:  -0.0111
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 562 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.160  0.860  0.955  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      1.67e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.144 1      1.070
## Year              1.144 16      1.004

```


Residuals from first author



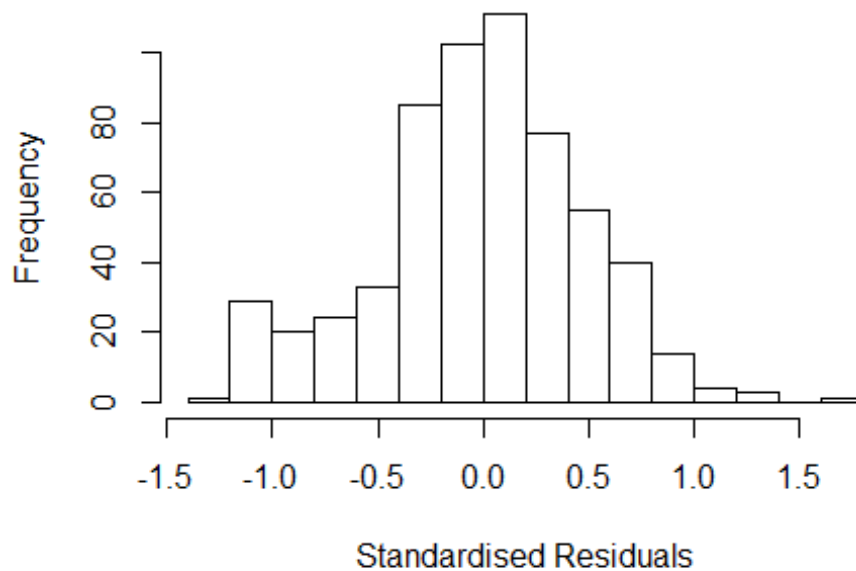
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23215 -0.29133  0.00474  0.29857  1.67815
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2456    0.1677    7.43  4e-13 ***
## FirstAuthorFemale1 -0.0134    0.0478   -0.28    0.78
## Year1997         -0.3354    0.3393   -0.99    0.32
## Year1998         -0.0660    0.2135   -0.31    0.76
## Year1999         -0.0633    0.1970   -0.32    0.75
## Year2000          0.0694    0.2609    0.27    0.79
## Year2001         -0.1890    0.2033   -0.93    0.35
## Year2002         -0.2007    0.2408   -0.83    0.40
## Year2003         -0.1399    0.2116   -0.66    0.51
## Year2004         -0.1598    0.2238   -0.71    0.48
## Year2005         -0.2143    0.1952   -1.10    0.27
## Year2006         -0.1064    0.1922   -0.55    0.58
```

```

## Year2007          -0.0729      0.1892   -0.39      0.70
## Year2008          -0.0773      0.1764   -0.44      0.66
## Year2009          -0.1237      0.1739   -0.71      0.48
## Year2010          -0.1490      0.1713   -0.87      0.38
## Year2011          -0.2209      0.1789   -1.23      0.22
## Year2012          -0.0736      0.1718   -0.43      0.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0191, Adjusted R-squared:  -0.00959
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 559 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.158  0.856  0.954  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      1.67e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



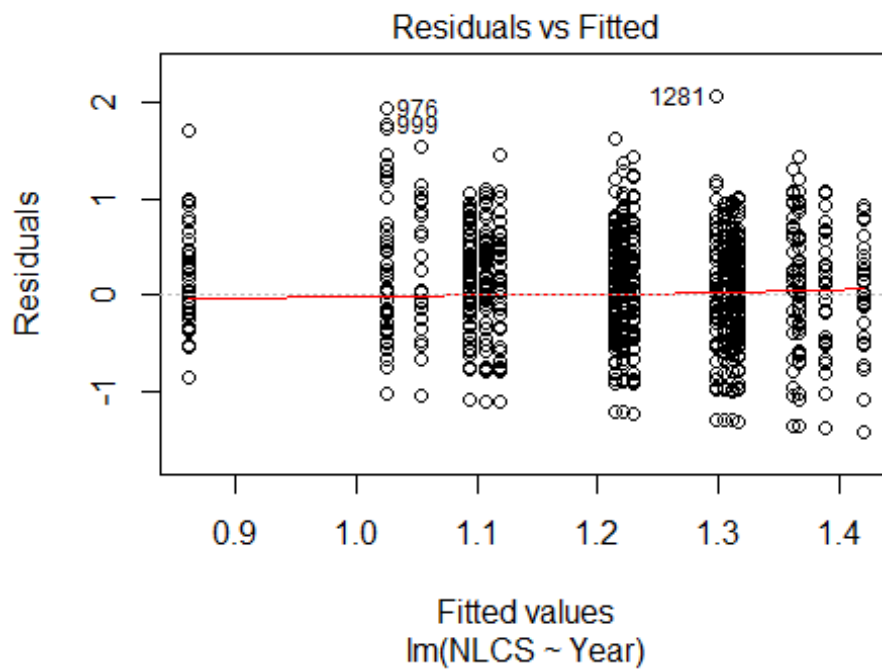
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24929 -0.29410 0.00747 0.30294 1.67506
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2493 0.1710 7.31 9e-13 ***
## LastAuthorFemale1 -0.0222 0.0567 -0.39 0.69
## Year1997 -0.3371 0.3379 -1.00 0.32
## Year1998 -0.0713 0.2152 -0.33 0.74
## Year1999 -0.0702 0.1985 -0.35 0.72
## Year2000 0.0629 0.2651 0.24 0.81
## Year2001 -0.1918 0.2042 -0.94 0.35
## Year2002 -0.2059 0.2423 -0.85 0.40
## Year2003 -0.1431 0.2127 -0.67 0.50
## Year2004 -0.1634 0.2250 -0.73 0.47
## Year2005 -0.2199 0.1996 -1.10 0.27
## Year2006 -0.1117 0.1960 -0.57 0.57
```

```

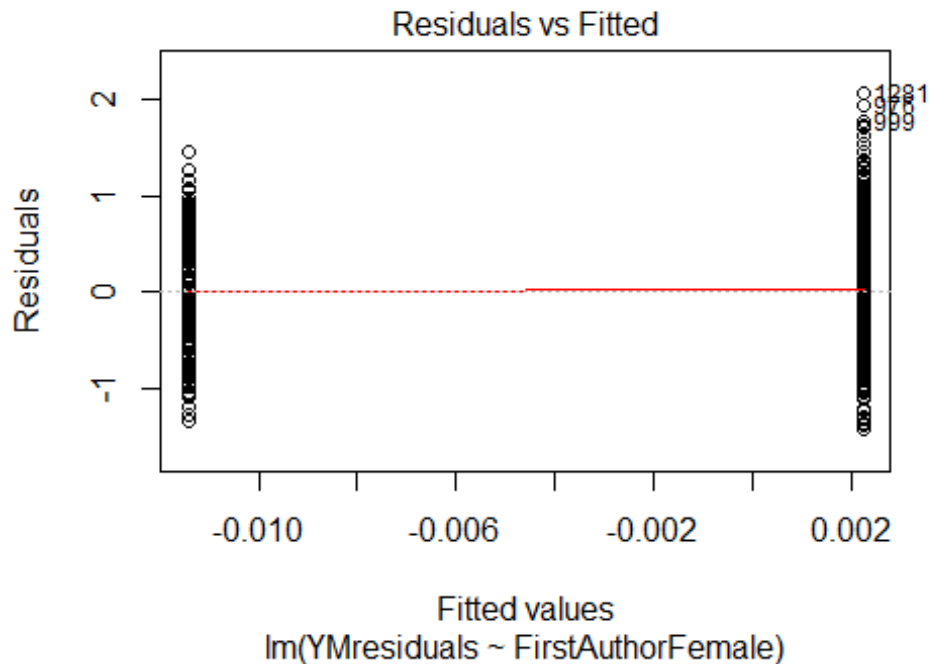
## Year2007          -0.0756      0.1905    -0.40      0.69
## Year2008          -0.0818      0.1785    -0.46      0.65
## Year2009          -0.1243      0.1744    -0.71      0.48
## Year2010          -0.1536      0.1743    -0.88      0.38
## Year2011          -0.2261      0.1810    -1.25      0.21
## Year2012          -0.0775      0.1741    -0.45      0.66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.0193, Adjusted R-squared:  -0.00936
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 561 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.164  0.863  0.955  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      1.67e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 599"
## [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
##  108  120   86  117  124  143  163  133  163  131  136  165  185  237  248
## 2011 2012
##  279  303
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32  43   31   33   43   38   58   58   53   53   65   71   72  103  118
## 2011 2012

```

```
## 124 123
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 40 26 27 38 34 52 52 43 41 55 59 58 85 101
## 2011 2012
## 96 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 = 47, df = 16, p-value = 6e-05
```

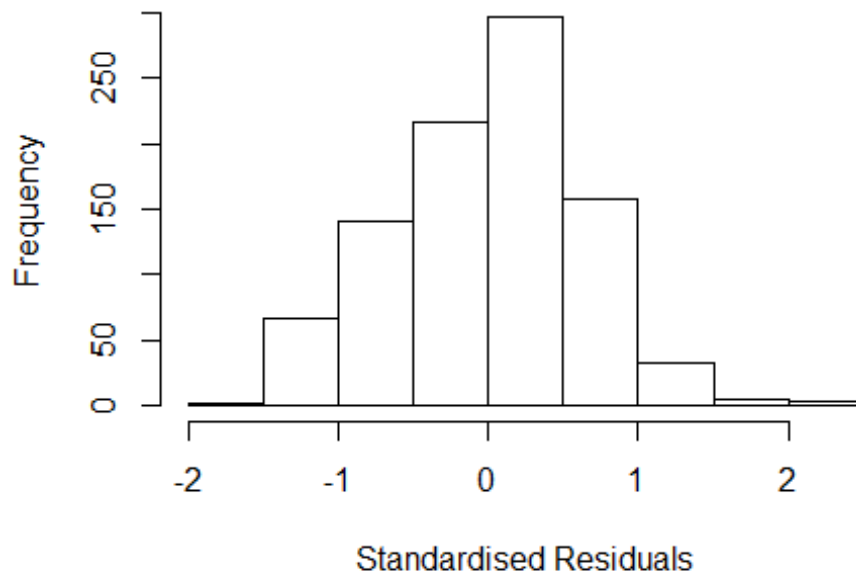


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.68, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.23067181887534"
## [1] "Male first author team size 2018 geometric mean: 2.55728388046922"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5000, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.02268576096372"
## [1] "Male last author team size 2018 geometric mean: 2.63011144082183"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, 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 1.188 1      1.090
## LastAuthorFemale  1.192 1      1.092
## UniqueAuthors    1.375 4      1.041
## Year              1.446 16     1.012
```

Residuals from first and last author and team size



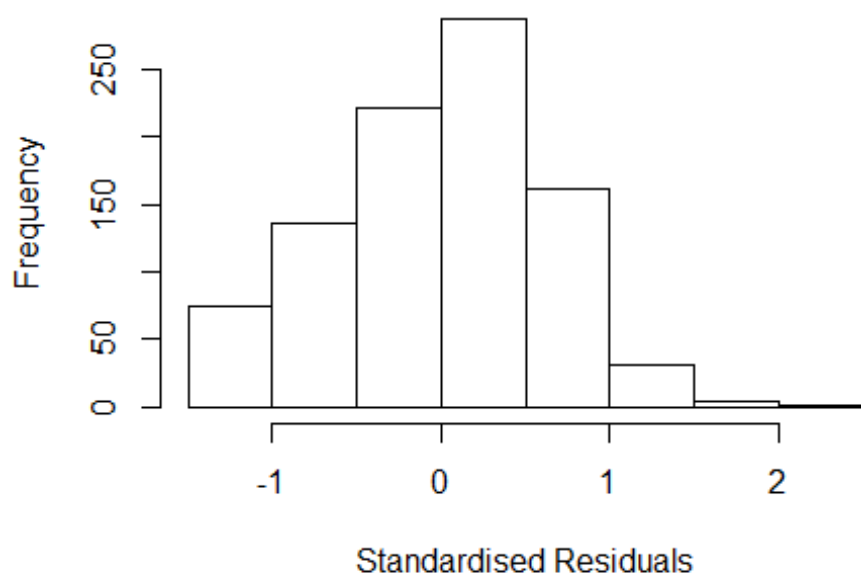
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5501 -0.4287 0.0643 0.4343 2.2057
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00327 0.17284 5.80 8.9e-09 ***
## FirstAuthorFemale1 0.00740 0.05742 0.13 0.89746
## LastAuthorFemale1 -0.04294 0.06915 -0.62 0.53481
## UniqueAuthors2 0.19666 0.05758 3.42 0.00067 ***
## UniqueAuthors3 0.23976 0.05970 4.02 6.4e-05 ***
## UniqueAuthors4 0.24421 0.08185 2.98 0.00292 **
## UniqueAuthors5 0.39003 0.09031 4.32 1.7e-05 ***
## Year1997 -0.03709 0.20426 -0.18 0.85595
## Year1998 0.29553 0.21586 1.37 0.17132
## Year1999 0.35020 0.21192 1.65 0.09878 .
```

```

## Year2000      0.22996      0.21540      1.07  0.28599
## Year2001      0.23192      0.25635      0.90  0.36587
## Year2002     -0.24196      0.21506     -1.13  0.26086
## Year2003      0.03520      0.20125      0.17  0.86119
## Year2004      0.08916      0.20272      0.44  0.66016
## Year2005     -0.31206      0.19536     -1.60  0.11054
## Year2006      0.00036      0.19471      0.00  0.99853
## Year2007      0.13105      0.18438      0.71  0.47742
## Year2008      0.21973      0.18751      1.17  0.24158
## Year2009      0.07491      0.18116      0.41  0.67935
## Year2010     -0.06257      0.17984     -0.35  0.72798
## Year2011      0.05984      0.18254      0.33  0.74311
## Year2012      0.12314      0.18055      0.68  0.49538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.636
## Multiple R-squared:  0.0811, Adjusted R-squared:  0.0585
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 847 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.205  0.872  0.950  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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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  1.168 1      1.081
## Year              1.084 16      1.003

```


Residuals from first and last author



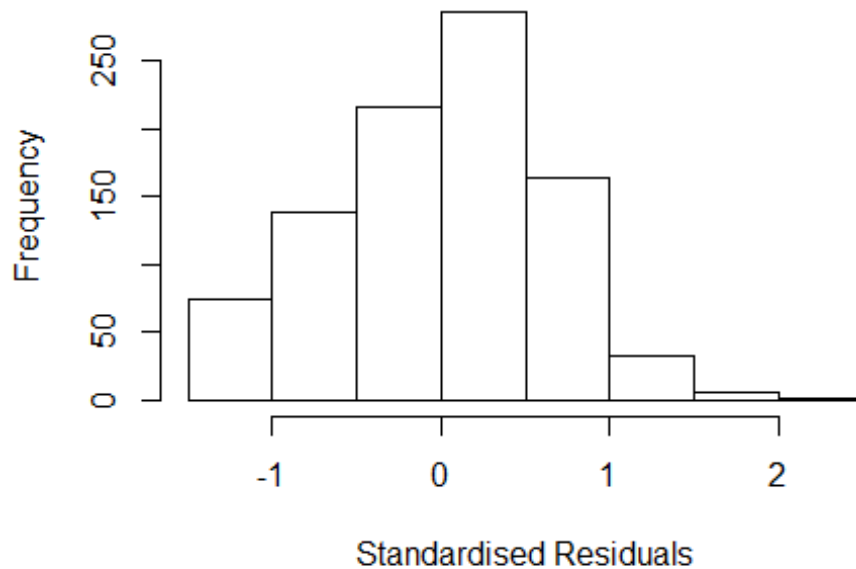
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.469 -0.469 0.055 0.418 2.102
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1061 0.1807 6.12 1.4e-09 ***
## FirstAuthorFemale1 0.0460 0.0575 0.80 0.42
## LastAuthorFemale1 -0.0673 0.0702 -0.96 0.34
## Year1997 -0.0346 0.2151 -0.16 0.87
## Year1998 0.3626 0.2241 1.62 0.11
## Year1999 0.3583 0.2219 1.61 0.11
## Year2000 0.2405 0.2254 1.07 0.29
## Year2001 0.2633 0.2633 1.00 0.32
## Year2002 -0.1866 0.2223 -0.84 0.40
## Year2003 0.0602 0.2089 0.29 0.77
## Year2004 0.1528 0.2120 0.72 0.47
## Year2005 -0.2765 0.2037 -1.36 0.17
```

```

## Year2006          0.0194      0.2023      0.10      0.92
## Year2007          0.1906      0.1941      0.98      0.33
## Year2008          0.2677      0.1960      1.37      0.17
## Year2009          0.1210      0.1907      0.63      0.53
## Year2010         -0.0111      0.1899     -0.06      0.95
## Year2011          0.1177      0.1920      0.61      0.54
## Year2012          0.1858      0.1886      0.99      0.32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.0516, Adjusted R-squared:  0.0327
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 849 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.267  0.865   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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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
## -1.4593 -0.4662 0.0599 0.4206 2.1030
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1069 0.1816 6.09 1.6e-09 ***
## FirstAuthorFemale1 0.0252 0.0551 0.46 0.65
## Year1997 -0.0365 0.2159 -0.17 0.87
## Year1998 0.3524 0.2248 1.57 0.12
## Year1999 0.3502 0.2224 1.57 0.12
## Year2000 0.2346 0.2260 1.04 0.30
## Year2001 0.2541 0.2652 0.96 0.34
## Year2002 -0.1874 0.2222 -0.84 0.40
## Year2003 0.0584 0.2098 0.28 0.78
## Year2004 0.1511 0.2130 0.71 0.48
## Year2005 -0.2857 0.2051 -1.39 0.16
## Year2006 0.0133 0.2046 0.07 0.95
```

```

## Year2007          0.1834      0.1954      0.94      0.35
## Year2008          0.2626      0.1974      1.33      0.18
## Year2009          0.1155      0.1917      0.60      0.55
## Year2010         -0.0200      0.1910     -0.10      0.92
## Year2011          0.1102      0.1932      0.57      0.57
## Year2012          0.1808      0.1898      0.95      0.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.0505, Adjusted R-squared:  0.0326
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 847 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.864  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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.052 1      1.026
## Year      1.052 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
## -1.4682 -0.4648  0.0553  0.4254  2.0925

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1141    0.1817   6.13 1.3e-09 ***
## LastAuthorFemale1 -0.0485    0.0663  -0.73   0.46
## Year1997         -0.0406    0.2162  -0.19   0.85
## Year1998          0.3542    0.2248   1.58   0.12
## Year1999          0.3522    0.2228   1.58   0.11
## Year2000          0.2381    0.2258   1.05   0.29
## Year2001          0.2608    0.2634   0.99   0.32
## Year2002         -0.1892    0.2233  -0.85   0.40
## Year2003          0.0591    0.2096   0.28   0.78
## Year2004          0.1544    0.2127   0.73   0.47
## Year2005         -0.2745    0.2041  -1.35   0.18
## Year2006          0.0178    0.2032   0.09   0.93
## Year2007          0.1899    0.1951   0.97   0.33
## Year2008          0.2677    0.1968   1.36   0.17
## Year2009          0.1182    0.1918   0.62   0.54
## Year2010         -0.0138    0.1910  -0.07   0.94
## Year2011          0.1148    0.1931   0.59   0.55
## Year2012          0.1859    0.1896   0.98   0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.644
## Multiple R-squared:  0.051, Adjusted R-squared:  0.0331
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 848 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.866  0.947  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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 919"
## [1] ""

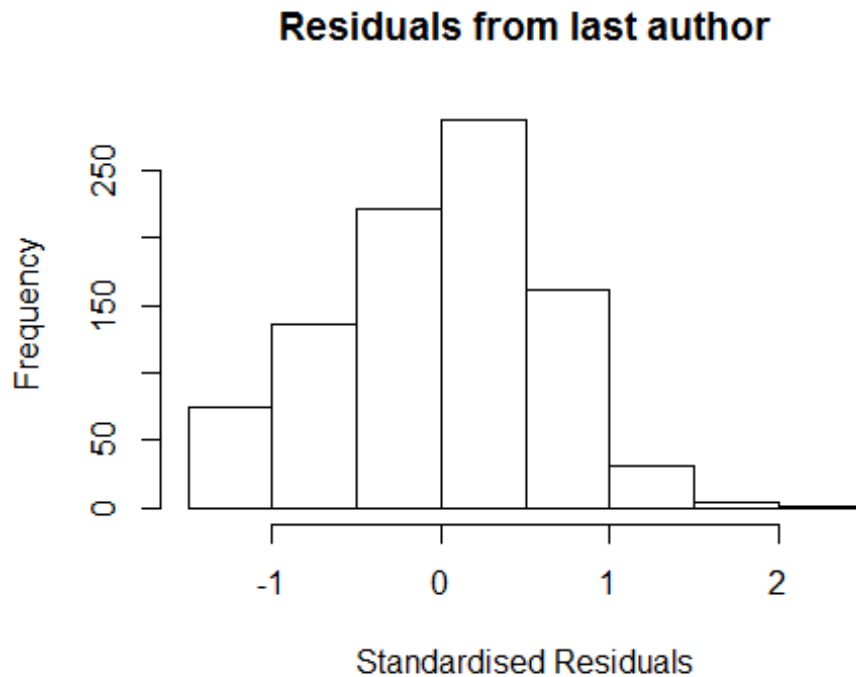
```

```

## [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
## 34 23 43 27 35 37 37 33 41 34 39 34 34 20 10
## 2011 2012
## 25 26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 1 7 4 3 8 4 9 10 10 15 18 13 11 4
## 2011 2012
## 14 9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 1 6 3 3 7 4 8 10 6 10 11 9 9 4
## 2011 2012
## 11 9
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3"
## [1] "Male first author team size 2018 geometric mean: 2.98894947601881"

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

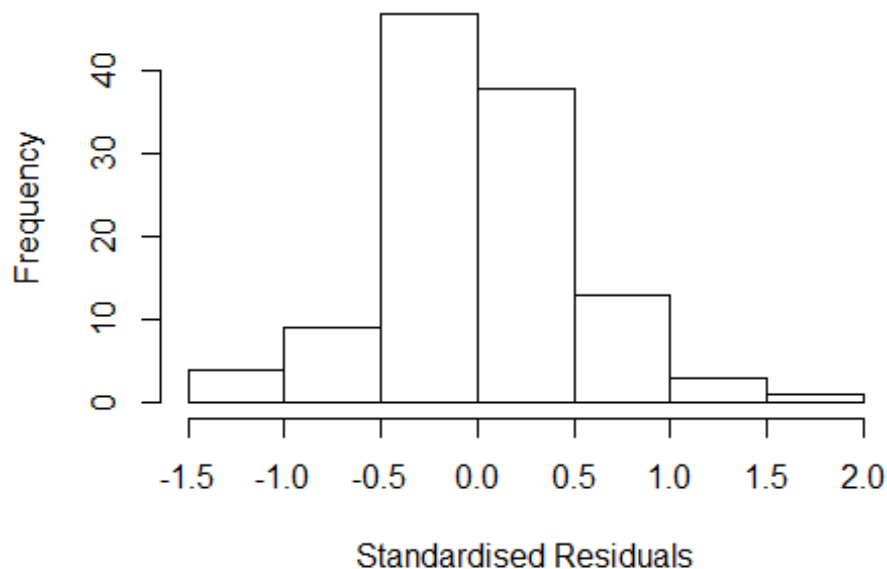
```



```
##
## 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] "Female last author team size 2018 geometric mean: NaN"
## [1] "Male last author team size 2018 geometric mean: 2.98986879625213"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	3.335	1	1.826
LastAuthorFemale	2.601	1	1.613
UniqueAuthors	223.502	4	1.966
Year	814.073	16	1.233

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
## -1.17481 -0.27494 -0.00578  0.28764  1.71439
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.102      0.326   0.31  0.75383
## FirstAuthorFemale1  0.370      0.164   2.25  0.02665 *
## LastAuthorFemale1 -0.570      0.151  -3.76  0.00029 ***
## UniqueAuthors2     0.252      0.173   1.46  0.14744
## UniqueAuthors3     0.254      0.186   1.36  0.17615
## UniqueAuthors4     0.148      0.229   0.65  0.52002
## UniqueAuthors5     0.395      0.206   1.92  0.05851 .
## Year1997           1.806      0.326   5.54  2.8e-07 ***
## Year1998           1.149      0.718   1.60  0.11306
## Year1999           0.976      0.537   1.82  0.07251 .
```

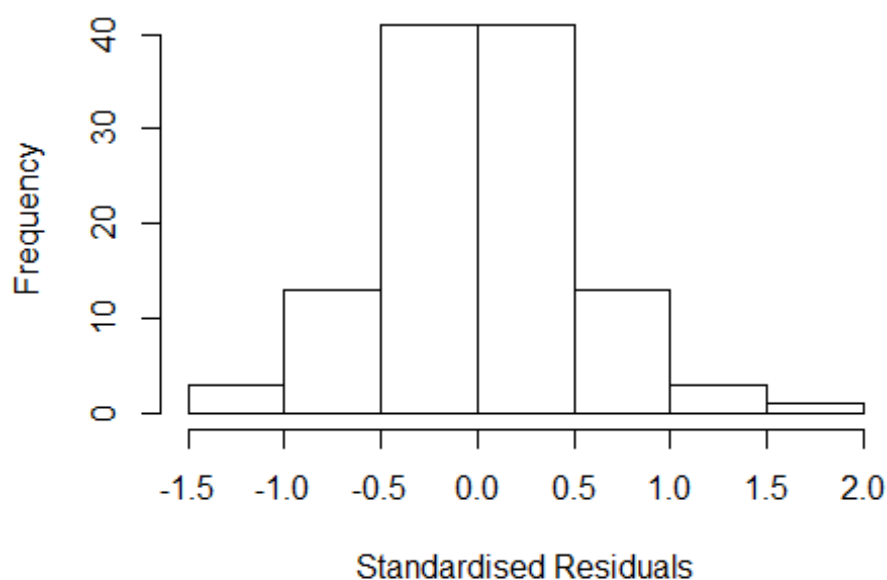


```

## Year2000          0.914      0.399      2.29  0.02413 *
## Year2001          1.099      0.542      2.03  0.04558 *
## Year2002          1.368      0.520      2.63  0.01006 *
## Year2003          0.841      0.385      2.18  0.03152 *
## Year2004          0.706      0.372      1.90  0.06116 .
## Year2005          0.678      0.388      1.75  0.08368 .
## Year2006          0.785      0.380      2.07  0.04148 *
## Year2007          0.836      0.387      2.16  0.03320 *
## Year2008          0.739      0.367      2.02  0.04674 *
## Year2009          1.149      0.394      2.92  0.00446 **
## Year2010          0.973      0.407      2.39  0.01895 *
## Year2011          0.819      0.384      2.13  0.03565 *
## Year2012          0.650      0.391      1.66  0.09985 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.492
## Multiple R-squared:  0.332, Adjusted R-squared:  0.172
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 107 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.200  0.894  0.964  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          8.70e-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 2.572 1          1.604
## LastAuthorFemale  1.887 1          1.374
## Year              4.484 16          1.048

```

Residuals from first and last author



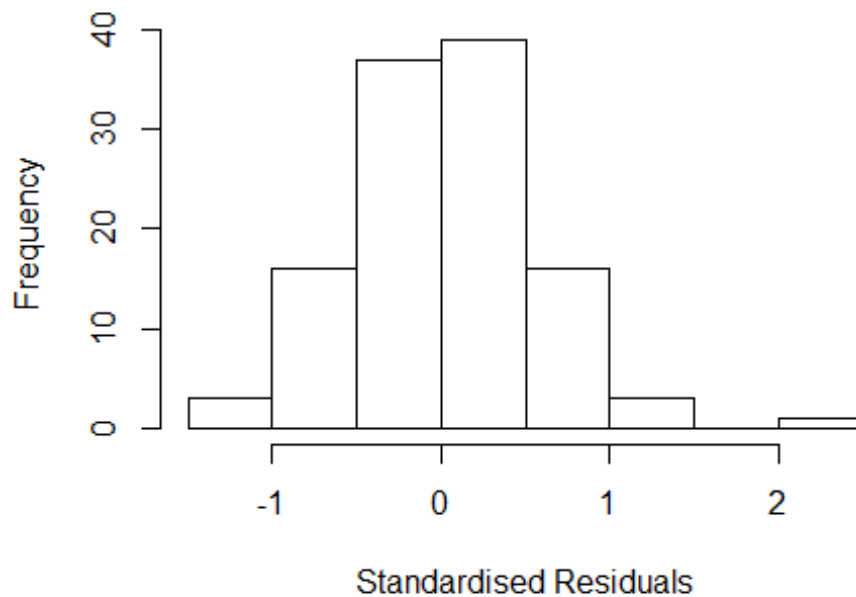
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1469 -0.3340 0.0211 0.3417 1.9258
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.143 0.259 0.55 0.58208
## FirstAuthorFemale1 0.407 0.148 2.76 0.00695 **
## LastAuthorFemale1 -0.506 0.143 -3.55 0.00060 ***
## Year1997 1.765 0.259 6.80 8.8e-10 ***
## Year1998 1.353 0.491 2.76 0.00697 **
## Year1999 0.984 0.411 2.40 0.01844 *
## Year2000 1.126 0.263 4.28 4.4e-05 ***
## Year2001 1.296 0.385 3.37 0.00110 **
## Year2002 1.520 0.419 3.63 0.00046 ***
## Year2003 0.968 0.307 3.16 0.00213 **
## Year2004 0.827 0.276 3.00 0.00348 **
## Year2005 0.723 0.324 2.24 0.02772 *
```

```

## Year2006          0.913      0.286      3.19  0.00190 **
## Year2007          1.009      0.313      3.23  0.00170 **
## Year2008          0.851      0.258      3.30  0.00137 **
## Year2009          1.297      0.297      4.36  3.2e-05 ***
## Year2010          1.161      0.291      3.99  0.00013 ***
## Year2011          1.004      0.289      3.47  0.00078 ***
## Year2012          0.778      0.326      2.39  0.01889 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.298, Adjusted R-squared:  0.167
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 102 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.171 0.882 0.957 0.913 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      8.70e-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 1.683 1      1.297
## Year              1.683 16      1.016

```

Residuals from first author



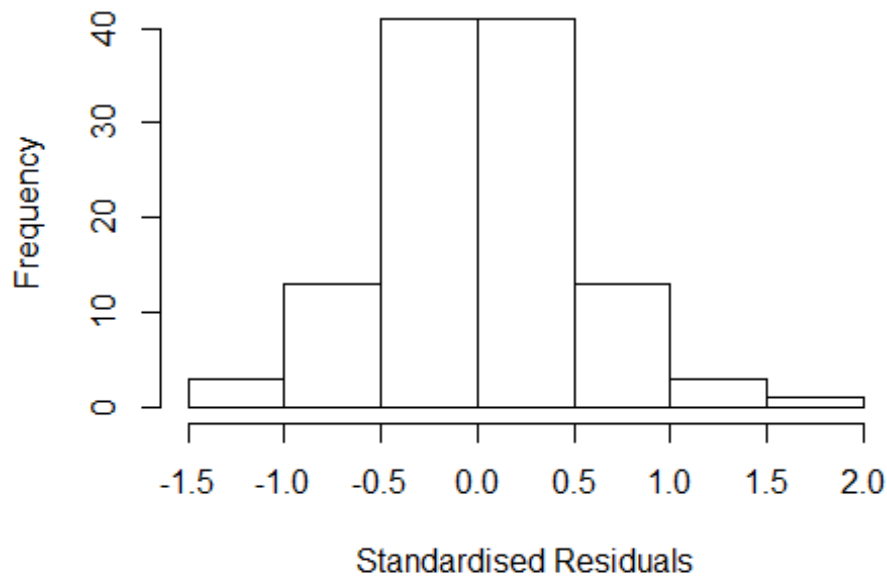
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09374 -0.35651  0.00533  0.39340  2.07433
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.00533    0.29385   -0.02  0.98557
## FirstAuthorFemale1  0.40670    0.15379    2.64  0.00954 **
## Year1997        1.91333    0.29385    6.51 3.3e-09 ***
## Year1998        1.44951    0.58978    2.46  0.01576 *
## Year1999        1.13028    0.42903    2.63  0.00981 **
## Year2000        1.27465    0.29707    4.29 4.2e-05 ***
## Year2001        1.44207    0.40263    3.58 0.00054 ***
## Year2002        1.51553    0.46994    3.22 0.00172 **
## Year2003        1.11371    0.33985    3.28 0.00146 **
## Year2004        0.92685    0.32461    2.86 0.00526 **
## Year2005        0.76981    0.35043    2.20 0.03042 *
## Year2006        1.00514    0.37653    2.67 0.00891 **
```

```

## Year2007          1.06440      0.35055      3.04  0.00308 **
## Year2008          0.94219      0.30539      3.09  0.00265 **
## Year2009          1.44871      0.32795      4.42  2.6e-05 ***
## Year2010          1.30955      0.32262      4.06  1.0e-04 ***
## Year2011          1.09414      0.33272      3.29  0.00140 **
## Year2012          0.92945      0.35268      2.64  0.00978 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.241, Adjusted R-squared:  0.108
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 100 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.180  0.895   0.955   0.918   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      8.70e-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 1.933 1          1.390
## Year              1.933 16          1.021

```

Residuals from last author



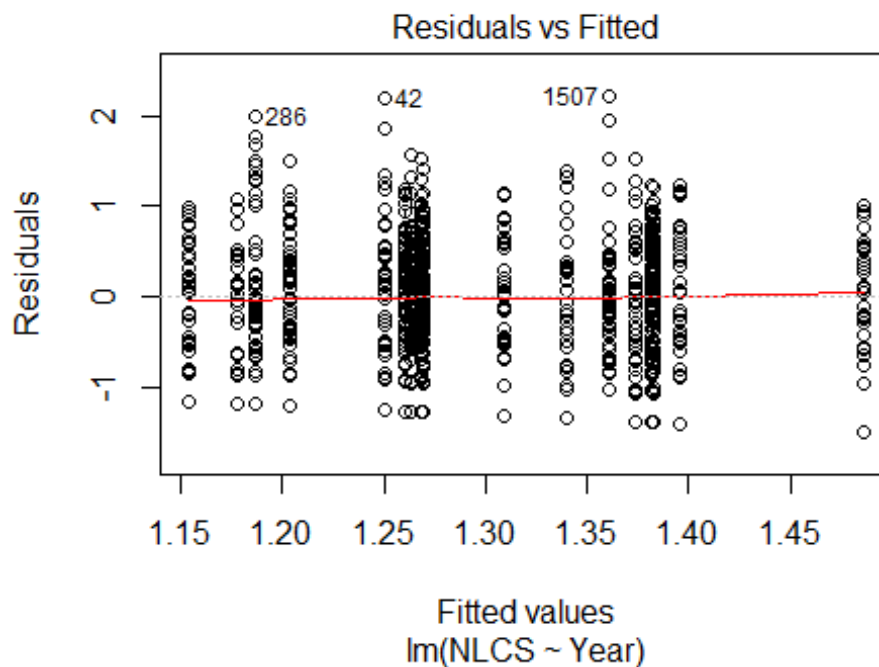
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.184 -0.349 0.021 0.358 1.723
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.346 0.395 0.88 0.38329
## LastAuthorFemale1 -0.512 0.146 -3.50 0.00070 ***
## Year1997 1.562 0.395 3.95 0.00015 ***
## Year1998 1.155 0.569 2.03 0.04533 *
## Year1999 0.781 0.505 1.55 0.12524
## Year2000 0.923 0.398 2.32 0.02239 *
## Year2001 1.093 0.488 2.24 0.02730 *
## Year2002 1.319 0.504 2.62 0.01024 *
## Year2003 0.822 0.441 1.87 0.06517 .
## Year2004 0.625 0.403 1.55 0.12428
## Year2005 0.521 0.435 1.20 0.23397
## Year2006 0.746 0.424 1.76 0.08182 .
```

```

## Year2007          0.911      0.459      1.98  0.04997 *
## Year2008          0.696      0.402      1.73  0.08608 .
## Year2009          1.136      0.411      2.76  0.00690 **
## Year2010          0.958      0.417      2.30  0.02389 *
## Year2011          0.838      0.419      2.00  0.04829 *
## Year2012          0.575      0.443      1.30  0.19676
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.261, Adjusted R-squared:  0.132
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 103 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.292  0.893  0.955  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      8.70e-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: 115"
## [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
## 115 109 125 111 137 134 90 96 85 88 110 138 150 166 161
## 2011 2012
## 146 173
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 24 53 42 52 33 31 40 35 32 53 69 76 90 85
## 2011 2012

```

```
##      83    92
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   37   23   41   35   42   19   24   34   32   26   40   62   60   76   64
## 2011 2012
##   69   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 = 32, df = 16, p-value = 0.01
```

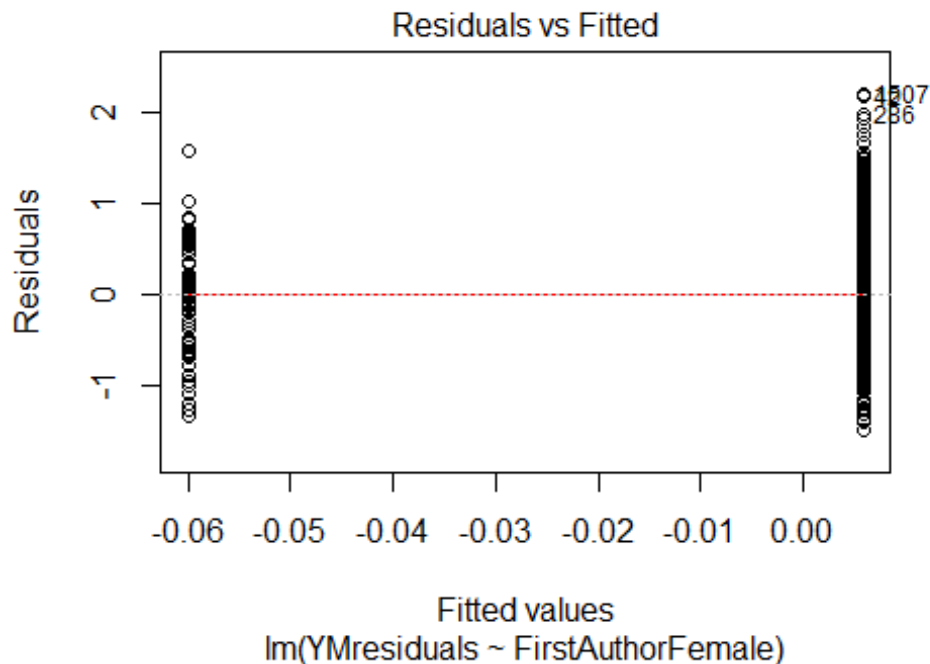


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.77, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 3.30192724889463"
## [1] "Male first author team size 2018 geometric mean: 2.79521148440675"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



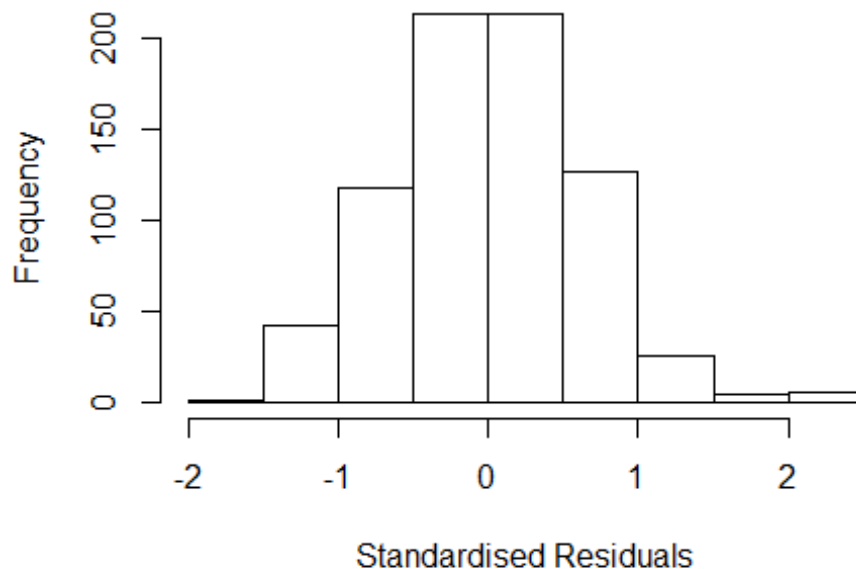
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2"
## [1] "Male last author team size 2018 geometric mean: 2.87226576410667"

## 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 = 13, 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 1.162  1      1.078
## LastAuthorFemale  1.211  1      1.100
## UniqueAuthors    1.493  4      1.051
## Year              1.757 16      1.018
```

Residuals from first and last author and team size



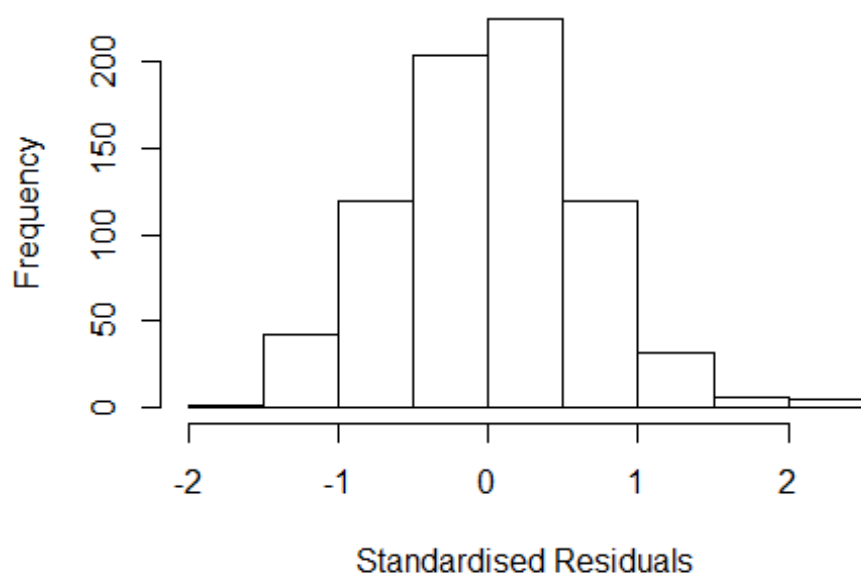
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.54007 -0.41435 0.00146 0.43382 2.42740
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00860 0.15756 6.40 2.8e-10 ***
## FirstAuthorFemale1 0.00340 0.08104 0.04 0.9666
## LastAuthorFemale1 -0.17339 0.10326 -1.68 0.0936 .
## UniqueAuthors2 0.20108 0.05908 3.40 0.0007 ***
## UniqueAuthors3 0.32186 0.06505 4.95 9.3e-07 ***
## UniqueAuthors4 0.29343 0.10880 2.70 0.0072 **
## UniqueAuthors5 0.10221 0.13823 0.74 0.4599
## Year1997 0.03000 0.20074 0.15 0.8812
## Year1998 -0.07414 0.20434 -0.36 0.7168
## Year1999 0.08442 0.19466 0.43 0.6647
```

```

## Year2000      0.00303      0.18691      0.02      0.9871
## Year2001      0.27700      0.23495      1.18      0.2388
## Year2002      0.36367      0.21181      1.72      0.0864 .
## Year2003      0.21741      0.18898      1.15      0.2503
## Year2004      0.27169      0.21915      1.24      0.2155
## Year2005      0.11188      0.20860      0.54      0.5919
## Year2006      0.08451      0.17410      0.49      0.6275
## Year2007      0.09138      0.17321      0.53      0.5980
## Year2008      0.06677      0.17627      0.38      0.7049
## Year2009      0.09492      0.17382      0.55      0.5852
## Year2010      0.17170      0.17371      0.99      0.3233
## Year2011      0.20960      0.17339      1.21      0.2271
## Year2012      0.06130      0.17371      0.35      0.7243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.0718, Adjusted R-squared:  0.0438
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 697 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.082  0.879   0.948   0.908   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      1.33e-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 1.180 1      1.086
## LastAuthorFemale  1.221 1      1.105
## Year              1.188 16      1.005

```

Residuals from first and last author



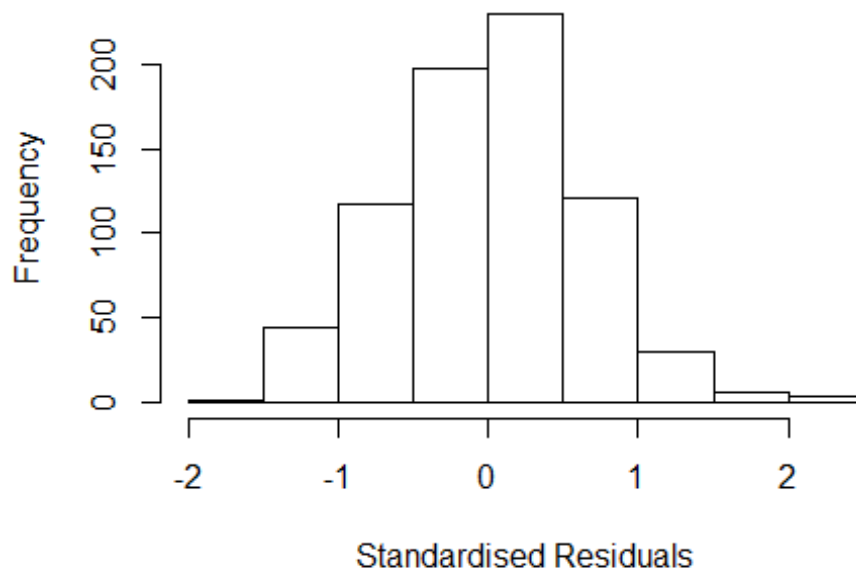
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5525 -0.4251 0.0254 0.4234 2.3195
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11649 0.15175 7.36 5e-13 ***
## FirstAuthorFemale1 0.02760 0.07947 0.35 0.728
## LastAuthorFemale1 -0.18965 0.10022 -1.89 0.059 .
## Year1997 0.03229 0.19720 0.16 0.870
## Year1998 -0.01965 0.20642 -0.10 0.924
## Year1999 0.09801 0.19426 0.50 0.614
## Year2000 0.00361 0.18395 0.02 0.984
## Year2001 0.33462 0.23368 1.43 0.153
## Year2002 0.43601 0.21256 2.05 0.041 *
## Year2003 0.26677 0.18863 1.41 0.158
## Year2004 0.33102 0.21684 1.53 0.127
## Year2005 0.14798 0.20196 0.73 0.464
```

```

## Year2006          0.18107      0.17072      1.06      0.289
## Year2007          0.13273      0.16780      0.79      0.429
## Year2008          0.15554      0.17102      0.91      0.363
## Year2009          0.15046      0.16822      0.89      0.371
## Year2010          0.26313      0.16830      1.56      0.118
## Year2011          0.31408      0.16680      1.88      0.060 .
## Year2012          0.13634      0.16729      0.82      0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0363, Adjusted R-squared:  0.0126
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 680 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.130  0.867  0.946  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      1.33e-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 1.071 1          1.035
## Year              1.071 16          1.002

```

Residuals from first author



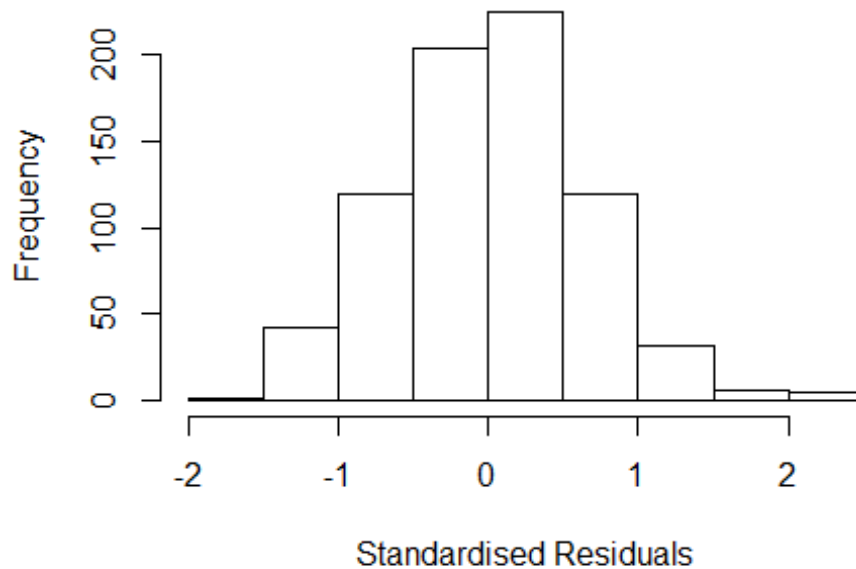
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5470 -0.4316 0.0324 0.4265 2.3264
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10965 0.14844 7.48 2.2e-13 ***
## FirstAuthorFemale1 -0.02183 0.07648 -0.29 0.775
## Year1997 0.03497 0.19232 0.18 0.856
## Year1998 -0.00497 0.20532 -0.02 0.981
## Year1999 0.09872 0.19118 0.52 0.606
## Year2000 0.01320 0.18121 0.07 0.942
## Year2001 0.33024 0.23178 1.42 0.155
## Year2002 0.43737 0.20738 2.11 0.035 *
## Year2003 0.27081 0.18643 1.45 0.147
## Year2004 0.32006 0.21445 1.49 0.136
## Year2005 0.15730 0.19924 0.79 0.430
## Year2006 0.18093 0.16953 1.07 0.286
```

```

## Year2007          0.11334    0.16297    0.70    0.487
## Year2008          0.15146    0.16781    0.90    0.367
## Year2009          0.15652    0.16485    0.95    0.343
## Year2010          0.27312    0.16439    1.66    0.097 .
## Year2011          0.29995    0.16242    1.85    0.065 .
## Year2012          0.12443    0.16445    0.76    0.449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.0307, Adjusted R-squared:  0.00826
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 677 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.136  0.872  0.948  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      1.33e-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 1.113 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
## -1.5519 -0.4283 0.0284 0.4213 2.3165
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11945 0.15169 7.38 4.3e-13 ***
## LastAuthorFemale1 -0.18000 0.09578 -1.88 0.061 .
## Year1997 0.03068 0.19711 0.16 0.876
## Year1998 -0.01824 0.20644 -0.09 0.930
## Year1999 0.09820 0.19419 0.51 0.613
## Year2000 0.00276 0.18433 0.01 0.988
## Year2001 0.33239 0.23310 1.43 0.154
## Year2002 0.43249 0.21232 2.04 0.042 *
## Year2003 0.26598 0.18916 1.41 0.160
## Year2004 0.32928 0.21464 1.53 0.125
## Year2005 0.14641 0.20210 0.72 0.469
## Year2006 0.17966 0.17107 1.05 0.294
```

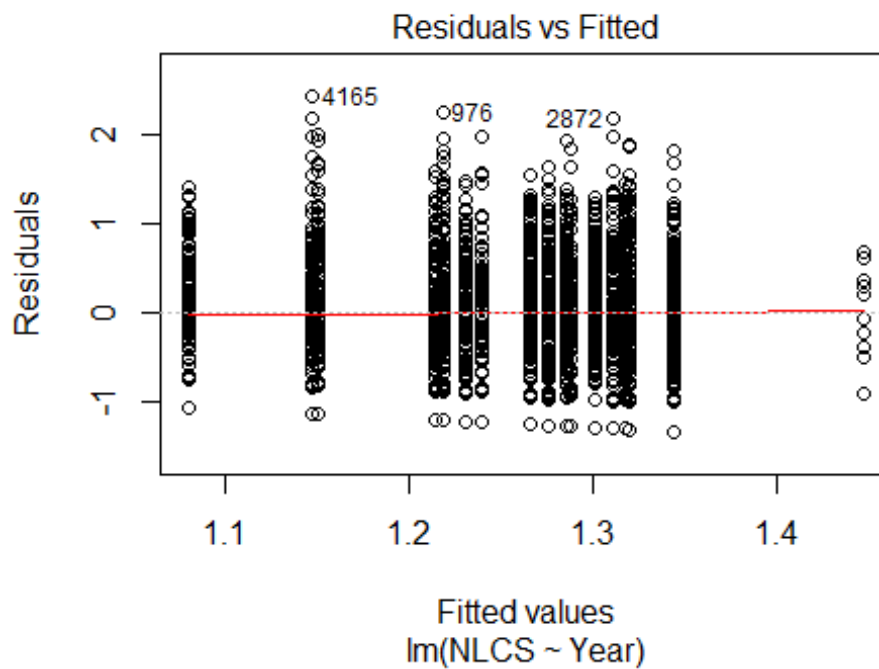


```

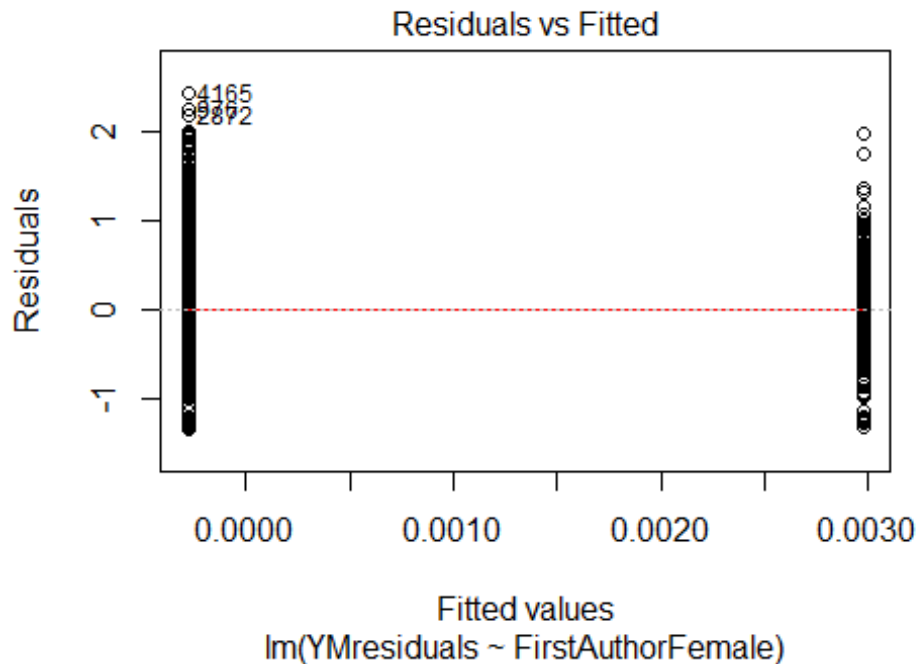
## Year2007      0.13041    0.16789    0.78    0.438
## Year2008      0.15548    0.17122    0.91    0.364
## Year2009      0.14913    0.16838    0.89    0.376
## Year2010      0.26307    0.16852    1.56    0.119
## Year2011      0.31324    0.16709    1.87    0.061 .
## Year2012      0.13495    0.16753    0.81    0.421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.0358, Adjusted R-squared:  0.0135
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 681 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.136  0.868  0.948  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      1.33e-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: 751"
## [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
##  335  412  393   29  347  265  275  281  271  300  375  412  423  477  478
## 2011 2012
##  460  521
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  101  136  143   10  130   82  113  127  118  139  154  189  197  220  237
## 2011 2012

```

```
## 210 238
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 109 117 7 107 55 95 96 96 118 120 154 148 175 176
## 2011 2012
## 168 176
## [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
```

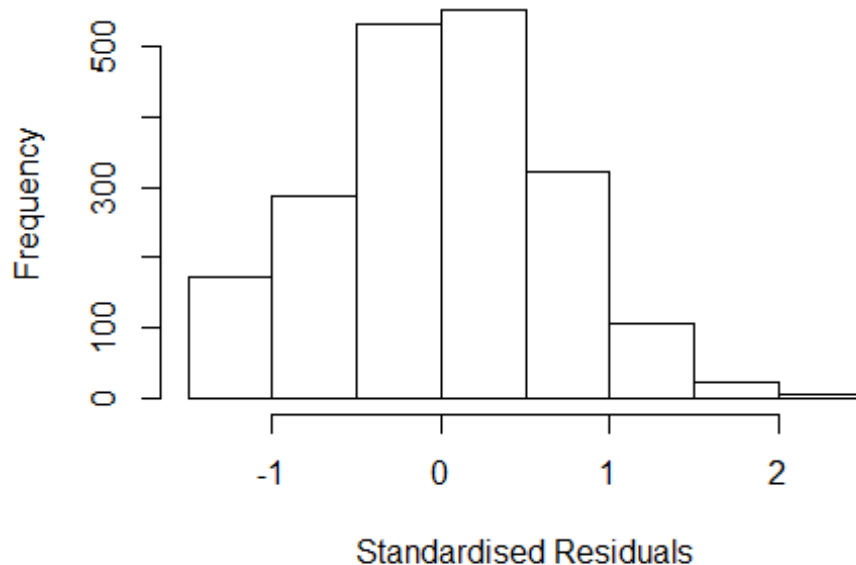


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 2.86687075329982"
## [1] "Male first author team size 2018 geometric mean: 2.90337039459286"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 480, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.98502160122875"
## [1] "Male last author team size 2018 geometric mean: 2.89085568309737"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 420, 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 1.104 1      1.051
## LastAuthorFemale  1.120 1      1.058
## UniqueAuthors     1.229 4      1.026
## Year              1.297 16     1.008
```

Residuals from first and last author and team size



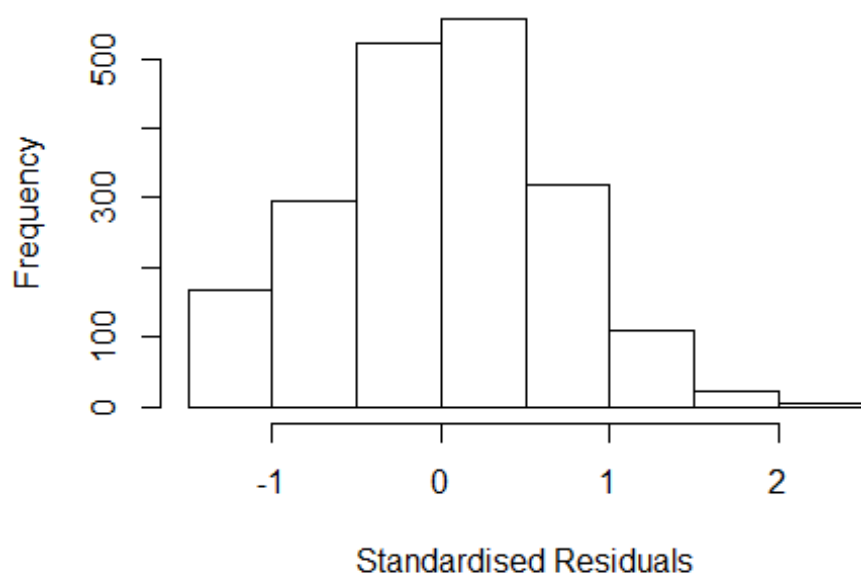
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3642 -0.4457 0.0106 0.4527 2.4294
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2138 0.0841 14.44 <2e-16 ***
## FirstAuthorFemale1 -0.0171 0.0512 -0.33 0.7388
## LastAuthorFemale1 0.0115 0.0602 0.19 0.8489
## UniqueAuthors2 0.1148 0.0415 2.76 0.0058 **
## UniqueAuthors3 0.0991 0.0443 2.24 0.0254 *
## UniqueAuthors4 0.1073 0.0563 1.91 0.0567 .
## UniqueAuthors5 0.0759 0.0625 1.21 0.2249
## Year1997 -0.1627 0.1147 -1.42 0.1563
## Year1998 -0.0994 0.1250 -0.80 0.4264
## Year1999 0.3153 0.2400 1.31 0.1891
```

```

## Year2000          -0.0654      0.1082    -0.60    0.5461
## Year2001          -0.0552      0.1486    -0.37    0.7105
## Year2002          -0.2069      0.1131    -1.83    0.0675 .
## Year2003           0.0357      0.1116     0.32    0.7492
## Year2004           0.0314      0.1216     0.26    0.7965
## Year2005          -0.0914      0.1020    -0.90    0.3706
## Year2006          -0.1829      0.0998    -1.83    0.0671 .
## Year2007           0.0242      0.0926     0.26    0.7940
## Year2008          -0.0258      0.0968    -0.27    0.7901
## Year2009          -0.0481      0.0929    -0.52    0.6051
## Year2010          -0.0746      0.0949    -0.79    0.4317
## Year2011           0.0280      0.0958     0.29    0.7700
## Year2012           0.0246      0.0972     0.25    0.8003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.675
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.00749
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 173 weights are ~= 1. The remaining 1826 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.168  0.868  0.950  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          5.00e-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 1.098 1          1.048
## LastAuthorFemale  1.120 1          1.058
## Year              1.095 16          1.003

```

Residuals from first and last author



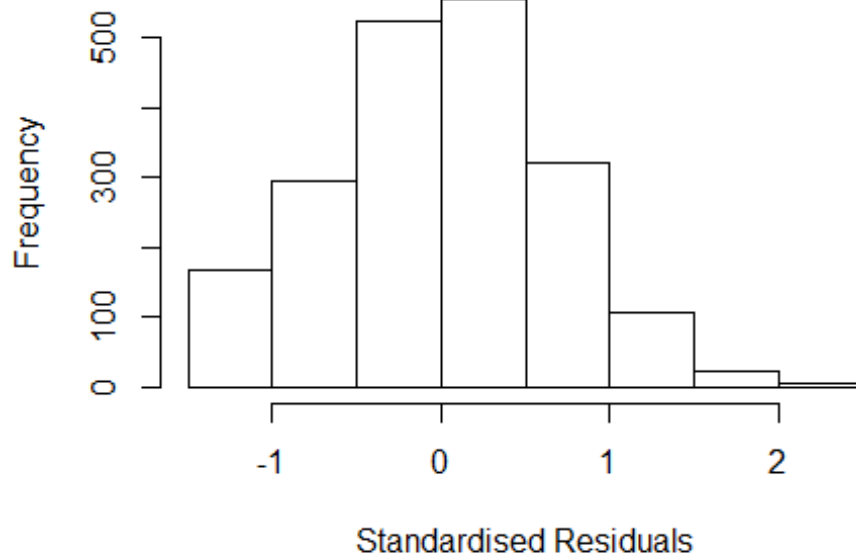
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.328 -0.442 0.014 0.458 2.469
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27159 0.08256 15.40 <2e-16 ***
## FirstAuthorFemale1 -0.00406 0.05054 -0.08 0.936
## LastAuthorFemale1 0.01473 0.05915 0.25 0.803
## Year1997 -0.16576 0.11496 -1.44 0.149
## Year1998 -0.09706 0.12627 -0.77 0.442
## Year1999 0.30756 0.23181 1.33 0.185
## Year2000 -0.06550 0.10907 -0.60 0.548
## Year2001 -0.03486 0.14639 -0.24 0.812
## Year2002 -0.18815 0.11300 -1.67 0.096 .
## Year2003 0.04729 0.11305 0.42 0.676
## Year2004 0.04182 0.12210 0.34 0.732
## Year2005 -0.06913 0.10227 -0.68 0.499
```

```

## Year2006      -0.16591    0.10014   -1.66    0.098 .
## Year2007      0.04158    0.09305    0.45    0.655
## Year2008     -0.00788    0.09697   -0.08    0.935
## Year2009     -0.03563    0.09346   -0.38    0.703
## Year2010     -0.05480    0.09494   -0.58    0.564
## Year2011      0.04853    0.09575    0.51    0.612
## Year2012      0.05047    0.09686    0.52    0.602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00458
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 167 weights are ~= 1. The remaining 1832 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.869  0.950  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      5.00e-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 1.039 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
## -1.3230 -0.4424 0.0151 0.4563 2.4685
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27237 0.08249 15.42 <2e-16 ***
## FirstAuthorFemale1 -0.00121 0.04927 -0.02 0.980
## Year1997 -0.16629 0.11498 -1.45 0.148
## Year1998 -0.09718 0.12629 -0.77 0.442
## Year1999 0.30851 0.23257 1.33 0.185
## Year2000 -0.06585 0.10901 -0.60 0.546
## Year2001 -0.03524 0.14639 -0.24 0.810
## Year2002 -0.18880 0.11297 -1.67 0.095 .
## Year2003 0.04778 0.11293 0.42 0.672
## Year2004 0.04294 0.12172 0.35 0.724
## Year2005 -0.06950 0.10221 -0.68 0.497
## Year2006 -0.16588 0.10012 -1.66 0.098 .
```

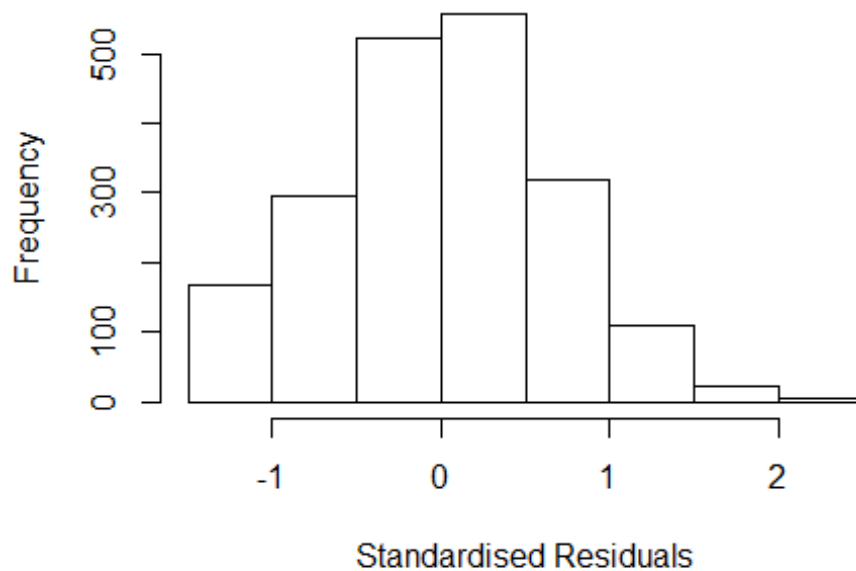


```

## Year2007          0.04196    0.09309    0.45    0.652
## Year2008          -0.00832    0.09694   -0.09    0.932
## Year2009          -0.03570    0.09346   -0.38    0.703
## Year2010          -0.05495    0.09495   -0.58    0.563
## Year2011           0.04911    0.09563    0.51    0.608
## Year2012           0.05060    0.09683    0.52    0.601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00506
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1829 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.869  0.950  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      5.00e-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 1.06 1          1.029
## Year              1.06 16          1.002

```

Residuals from last author



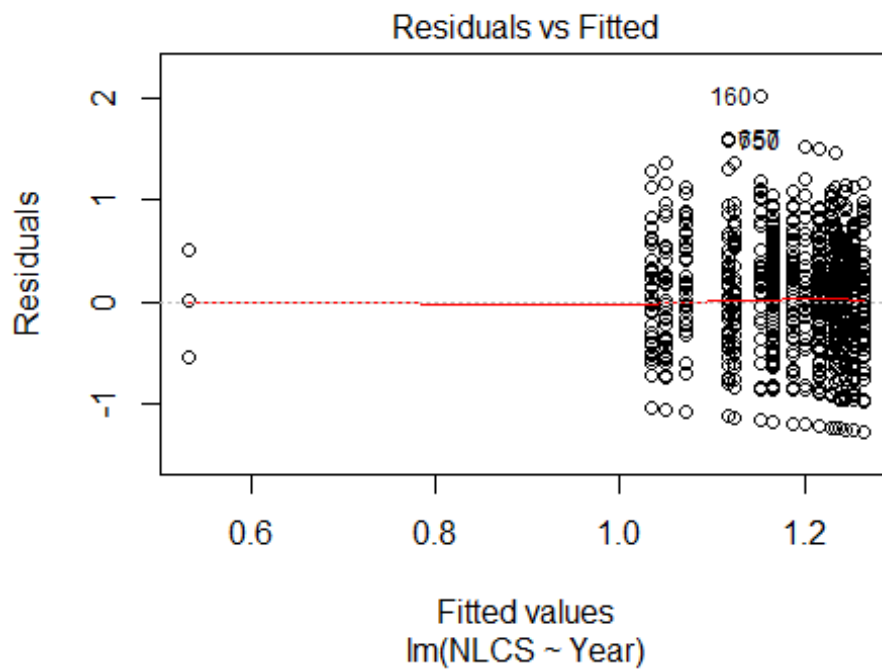
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3267 -0.4415 0.0136 0.4573 2.4695
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27115 0.08212 15.48 <2e-16 ***
## LastAuthorFemale1 0.01375 0.05753 0.24 0.811
## Year1997 -0.16550 0.11479 -1.44 0.150
## Year1998 -0.09715 0.12621 -0.77 0.442
## Year1999 0.30757 0.23161 1.33 0.184
## Year2000 -0.06528 0.10894 -0.60 0.549
## Year2001 -0.03471 0.14630 -0.24 0.813
## Year2002 -0.18808 0.11299 -1.66 0.096 .
## Year2003 0.04735 0.11297 0.42 0.675
## Year2004 0.04211 0.12190 0.35 0.730
## Year2005 -0.06900 0.10222 -0.67 0.500
## Year2006 -0.16562 0.09989 -1.66 0.097 .
```

```

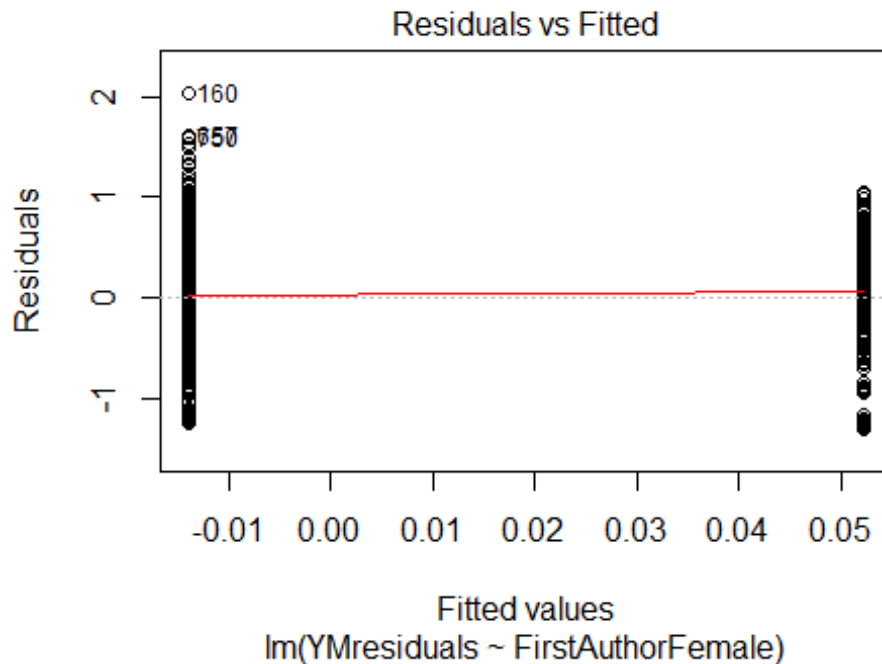
## Year2007          0.04184      0.09289      0.45      0.652
## Year2008          -0.00768      0.09684     -0.08      0.937
## Year2009          -0.03540      0.09330     -0.38      0.704
## Year2010          -0.05476      0.09491     -0.58      0.564
## Year2011           0.04856      0.09572      0.51      0.612
## Year2012           0.05055      0.09684      0.52      0.602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00508
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1831 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.151  0.869  0.950  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      5.00e-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: 1999"
## [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
##   10  115  150  124  158  167  135  115  129  122  151  157  192  213  206
## 2011 2012
##   224  225
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3   31   43   44   53   46   50   40   55   41   64   54   70   76   83
## 2011 2012

```

```
## 99 111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 27 36 39 48 42 42 36 45 34 54 43 55 63 75
## 2011 2012
## 76 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 = 20, df = 16, p-value = 0.2
```

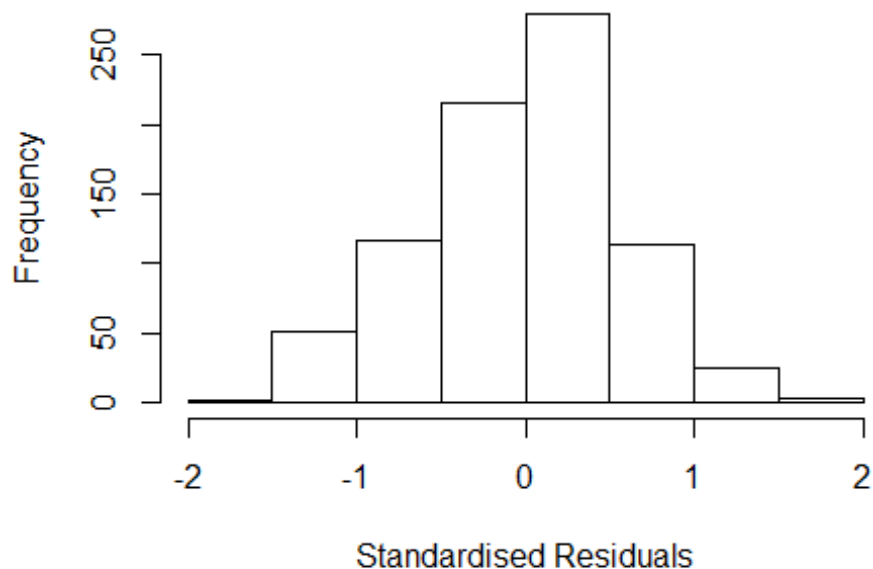


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.77, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.59310767413104"
## [1] "Male first author team size 2018 geometric mean: 2.93210431466958"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.98554487852364"
## [1] "Male last author team size 2018 geometric mean: 3.0867103617049"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, 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 1.230 1          1.109
## LastAuthorFemale  1.067 1          1.033
## UniqueAuthors    2.042 4          1.093
## Year             2.247 16          1.026
```

Residuals from first and last author and team size



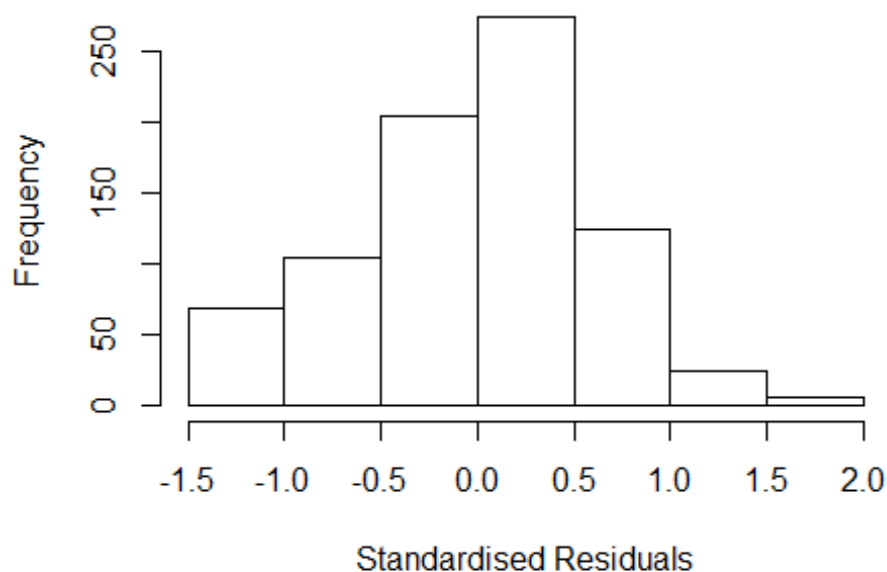
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5798 -0.4078 0.0362 0.3693 1.7618
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4431 0.1977 2.24 0.02526 *
## FirstAuthorFemale1 0.0854 0.0525 1.63 0.10448
## LastAuthorFemale1 -0.0449 0.0669 -0.67 0.50165
## UniqueAuthors2 0.1268 0.0654 1.94 0.05293 .
## UniqueAuthors3 0.2156 0.0717 3.01 0.00271 **
## UniqueAuthors4 0.4062 0.0842 4.82 1.7e-06 ***
## UniqueAuthors5 0.3223 0.0865 3.72 0.00021 ***
## Year1997 0.6921 0.2471 2.80 0.00522 **
## Year1998 0.4799 0.2584 1.86 0.06368 .
## Year1999 0.4742 0.2235 2.12 0.03422 *
```

```

## Year2000          0.5391      0.2193      2.46  0.01419 *
## Year2001          0.5170      0.2210      2.34  0.01955 *
## Year2002          0.6042      0.2225      2.72  0.00676 **
## Year2003          0.3967      0.2307      1.72  0.08593 .
## Year2004          0.6351      0.2226      2.85  0.00443 **
## Year2005          0.4188      0.2186      1.92  0.05580 .
## Year2006          0.7754      0.2104      3.69  0.00024 ***
## Year2007          0.5461      0.2167      2.52  0.01194 *
## Year2008          0.5907      0.2115      2.79  0.00535 **
## Year2009          0.6437      0.2195      2.93  0.00346 **
## Year2010          0.5944      0.2086      2.85  0.00449 **
## Year2011          0.5211      0.2095      2.49  0.01307 *
## Year2012          0.5058      0.2080      2.43  0.01526 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0736, Adjusted R-squared:  0.0475
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 730 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.310  0.847  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      1.24e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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  1.042 1      1.021
## Year              1.145 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
## -1.4369 -0.3998 0.0394 0.3890 1.6790
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5427 0.2765 1.96 0.0501 .
## FirstAuthorFemale1 0.1228 0.0521 2.36 0.0187 *
## LastAuthorFemale1 -0.0340 0.0682 -0.50 0.6186
## Year1997 0.6596 0.3116 2.12 0.0346 *
## Year1998 0.4903 0.3256 1.51 0.1325
## Year1999 0.4684 0.2983 1.57 0.1168
## Year2000 0.5806 0.2940 1.97 0.0487 *
## Year2001 0.5003 0.2964 1.69 0.0919 .
## Year2002 0.6289 0.2955 2.13 0.0336 *
## Year2003 0.4324 0.2988 1.45 0.1482
## Year2004 0.6914 0.2933 2.36 0.0187 *
## Year2005 0.4910 0.2916 1.68 0.0926 .
```

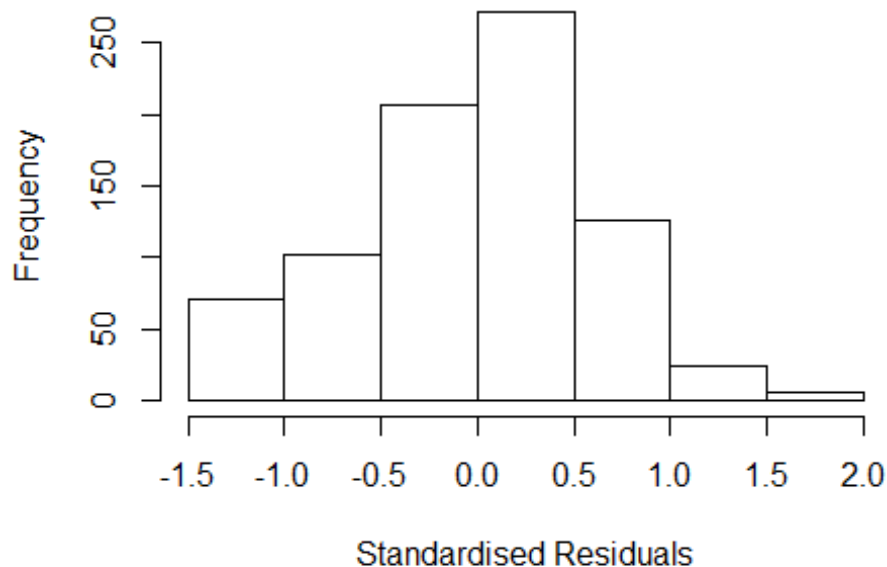


```

## Year2006          0.8054      0.2850      2.83      0.0048 **
## Year2007          0.5776      0.2896      1.99      0.0465 *
## Year2008          0.6650      0.2879      2.31      0.0211 *
## Year2009          0.7171      0.2936      2.44      0.0148 *
## Year2010          0.6711      0.2834      2.37      0.0181 *
## Year2011          0.6267      0.2843      2.20      0.0278 *
## Year2012          0.6112      0.2834      2.16      0.0313 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.575
## Multiple R-squared:  0.0343, Adjusted R-squared:  0.0122
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 741 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.375  0.856   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      1.24e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



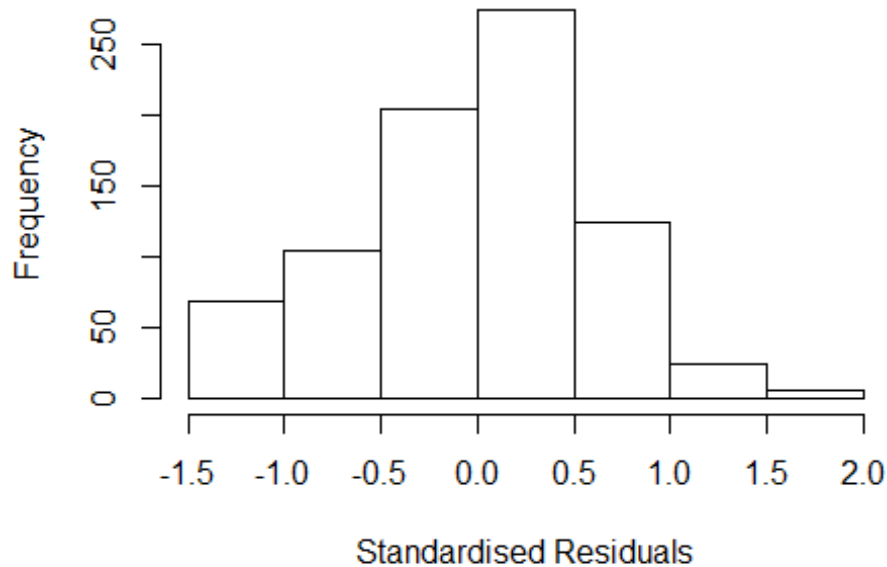
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4605 -0.4021 0.0437 0.3881 1.6817
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5323 0.2660 2.00 0.0457 *
## FirstAuthorFemale1 0.1150 0.0527 2.18 0.0295 *
## Year1997 0.6678 0.3025 2.21 0.0276 *
## Year1998 0.4998 0.3167 1.58 0.1150
## Year1999 0.4784 0.2883 1.66 0.0974 .
## Year2000 0.5868 0.2843 2.06 0.0393 *
## Year2001 0.5080 0.2871 1.77 0.0772 .
## Year2002 0.6356 0.2860 2.22 0.0265 *
## Year2003 0.4406 0.2891 1.52 0.1279
## Year2004 0.6984 0.2837 2.46 0.0141 *
## Year2005 0.5003 0.2818 1.78 0.0762 .
## Year2006 0.8133 0.2748 2.96 0.0032 **
```

```

## Year2007          0.5840      0.2799      2.09      0.0373 *
## Year2008          0.6701      0.2779      2.41      0.0161 *
## Year2009          0.7241      0.2839      2.55      0.0109 *
## Year2010          0.6774      0.2733      2.48      0.0134 *
## Year2011          0.6336      0.2744      2.31      0.0212 *
## Year2012          0.6180      0.2732      2.26      0.0240 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 737 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.378  0.858  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      1.24e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



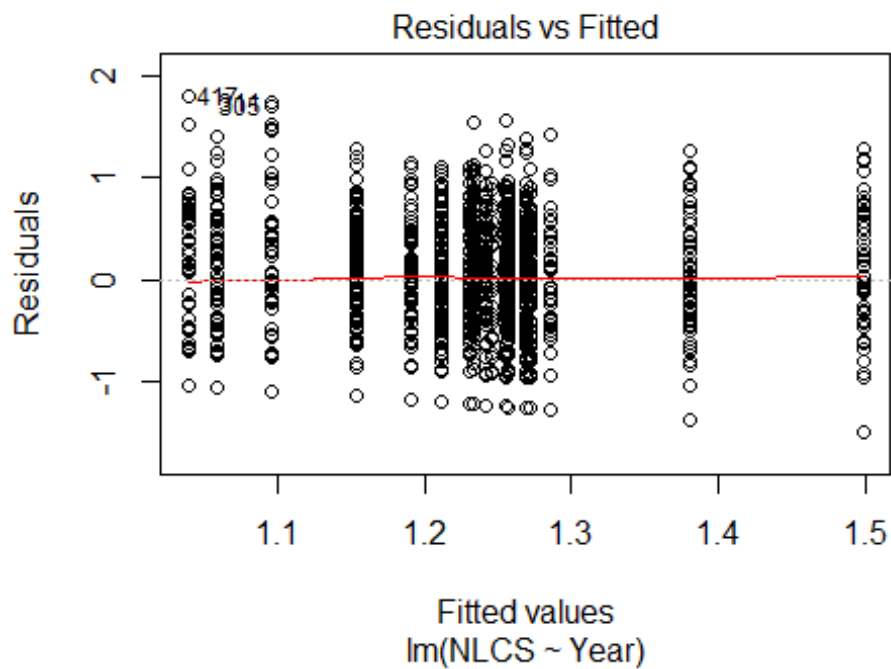
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3757 -0.4140 0.0401 0.3916 1.6586
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.53133 0.26533 2.00 0.0456 *
## LastAuthorFemale1 0.00304 0.06834 0.04 0.9645
## Year1997 0.67338 0.30082 2.24 0.0255 *
## Year1998 0.52393 0.31406 1.67 0.0957 .
## Year1999 0.48632 0.28863 1.68 0.0924 .
## Year2000 0.60826 0.28398 2.14 0.0325 *
## Year2001 0.53208 0.28454 1.87 0.0619 .
## Year2002 0.65555 0.28493 2.30 0.0217 *
## Year2003 0.47329 0.28815 1.64 0.1009
## Year2004 0.73537 0.28256 2.60 0.0094 **
## Year2005 0.53530 0.28142 1.90 0.0575 .
## Year2006 0.84130 0.27420 3.07 0.0022 **
```

```

## Year2007      0.60891    0.27790    2.19    0.0287 *
## Year2008      0.69125    0.27701    2.50    0.0128 *
## Year2009      0.73962    0.28357    2.61    0.0093 **
## Year2010      0.70485    0.27220    2.59    0.0098 **
## Year2011      0.66814    0.27309    2.45    0.0146 *
## Year2012      0.65756    0.27174    2.42    0.0158 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.58
## Multiple R-squared:  0.0275, Adjusted R-squared:  0.00654
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 735 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.394  0.853   0.945   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      1.24e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 806"
## [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
## 342 329 295 168 220 323 270 300 258 294 307 338 385 458 399
## 2011 2012
## 492 515
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 57 57 26 35 46 70 71 66 81 98 89 115 161 136
## 2011 2012

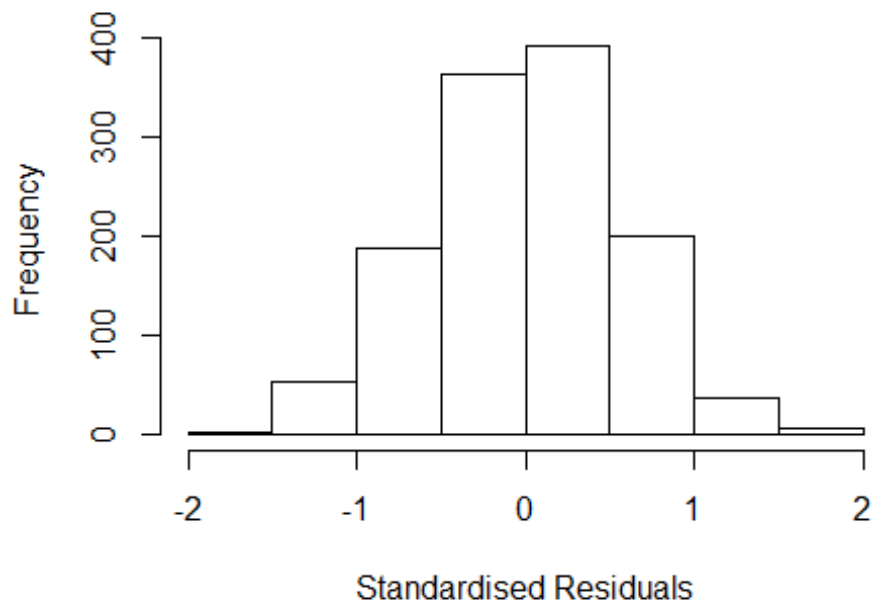
```

```
## 181 206
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 50 45 20 31 38 58 60 50 62 81 74 91 125 110
## 2011 2012
## 149 147
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.002
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.17, 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
## -1.6245 -0.3802  0.0189  0.3874  1.9120
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8850    0.1334   6.63 4.9e-11 ***
## FirstAuthorFemale1 0.0923    0.0464   1.99  0.047 *
## LastAuthorFemale1 0.0986    0.0588   1.68  0.094 .
## UniqueAuthors2    0.2025    0.0505   4.01 6.5e-05 ***
## UniqueAuthors3    0.3393    0.0501   6.77 2.0e-11 ***
## UniqueAuthors4    0.3529    0.0636   5.55 3.5e-08 ***
## UniqueAuthors5    0.4521    0.0630   7.18 1.2e-12 ***
## Year1997         -0.1005    0.1764  -0.57  0.569
## Year1998          0.3063    0.1545   1.98  0.048 *
## Year1999          0.1044    0.1667   0.63  0.531
```

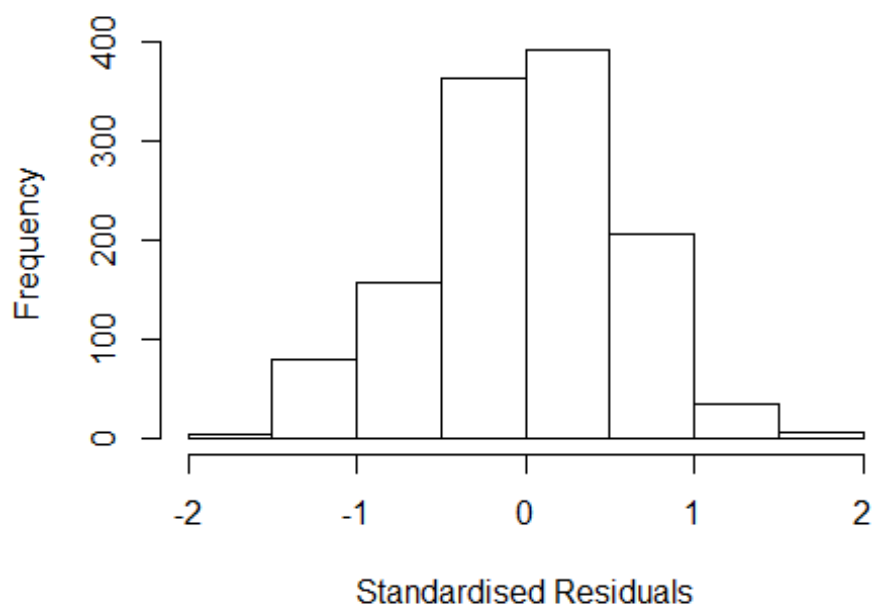


```

## Year2000          0.1446      0.1572      0.92      0.358
## Year2001          0.4002      0.1822      2.20      0.028 *
## Year2002         -0.0469      0.1599     -0.29      0.769
## Year2003          0.1540      0.1566      0.98      0.325
## Year2004          0.1061      0.1600      0.66      0.508
## Year2005          0.0667      0.1485      0.45      0.653
## Year2006          0.1484      0.1432      1.04      0.300
## Year2007          0.0800      0.1435      0.56      0.577
## Year2008          0.0888      0.1407      0.63      0.528
## Year2009          0.1564      0.1360      1.15      0.250
## Year2010          0.0142      0.1421      0.10      0.920
## Year2011          0.0856      0.1350      0.63      0.527
## Year2012          0.0624      0.1350      0.46      0.644
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.0984, Adjusted R-squared:  0.0821
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.234  0.864  0.950  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      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 1.065 1      1.032
## LastAuthorFemale  1.024 1      1.012
## Year              1.082 16      1.002

```

Residuals from first and last author



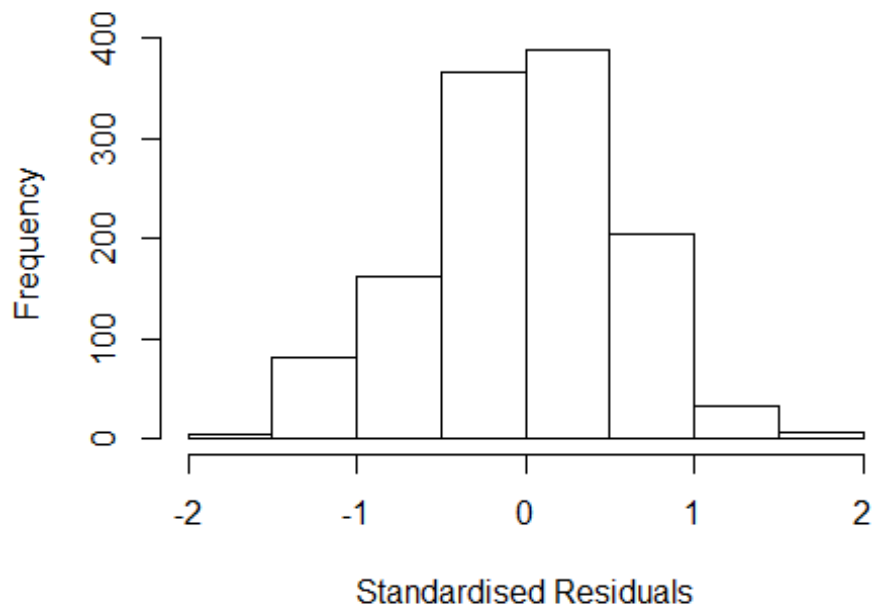
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5136 -0.4023 0.0121 0.3984 1.9181
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0010 0.1366 7.33 4.3e-13 ***
## FirstAuthorFemale1 0.1478 0.0467 3.16 0.0016 **
## LastAuthorFemale1 0.1177 0.0578 2.04 0.0419 *
## Year1997 -0.0740 0.1787 -0.41 0.6787
## Year1998 0.3516 0.1635 2.15 0.0317 *
## Year1999 0.1248 0.1807 0.69 0.4898
## Year2000 0.1889 0.1657 1.14 0.2545
## Year2001 0.5127 0.1820 2.82 0.0049 **
## Year2002 0.0153 0.1649 0.09 0.9263
## Year2003 0.1917 0.1595 1.20 0.2295
## Year2004 0.1870 0.1639 1.14 0.2539
## Year2005 0.1515 0.1552 0.98 0.3294
```

```

## Year2006          0.2463      0.1505      1.64      0.1020
## Year2007          0.2037      0.1527      1.33      0.1823
## Year2008          0.2190      0.1499      1.46      0.1443
## Year2009          0.2604      0.1442      1.81      0.0712 .
## Year2010          0.1446      0.1489      0.97      0.3316
## Year2011          0.2146      0.1431      1.50      0.1340
## Year2012          0.2130      0.1416      1.50      0.1328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0418, Adjusted R-squared:  0.0277
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1127 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.263  0.869  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      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 1.066 1      1.032
## Year              1.066 16      1.002

```

Residuals from first author

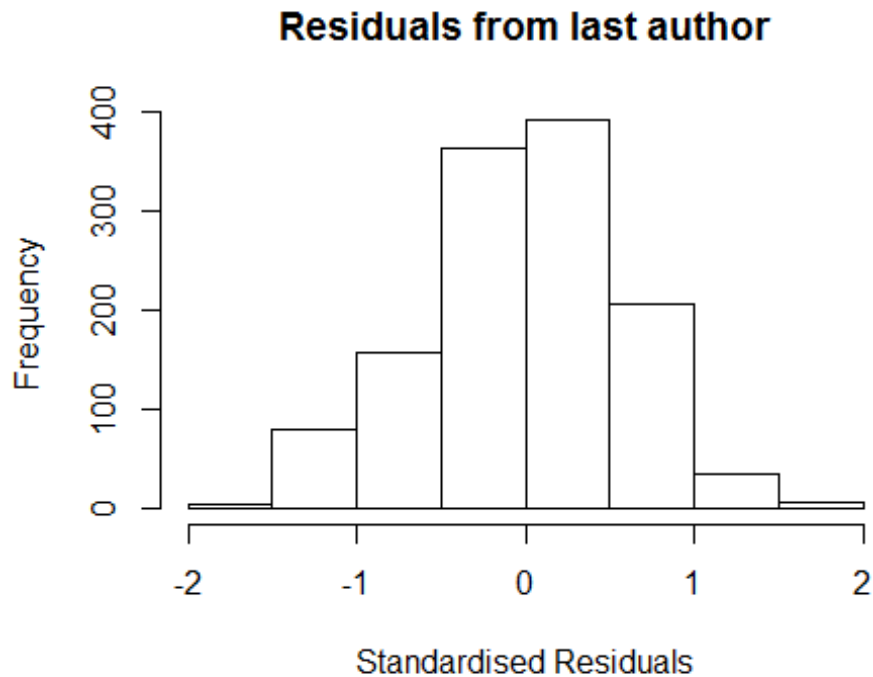


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.51896 -0.40569  0.00804  0.40475  1.90937
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0069    0.1372   7.34 3.9e-13 ***
## FirstAuthorFemale1  0.1592    0.0470   3.38 0.00074 ***
## Year1997        -0.0713    0.1802  -0.40 0.69249
## Year1998         0.3608    0.1646   2.19 0.02857 *
## Year1999         0.1232    0.1824   0.68 0.49951
## Year2000         0.1821    0.1661   1.10 0.27301
## Year2001         0.5120    0.1814   2.82 0.00484 **
## Year2002         0.0181    0.1662   0.11 0.91322
## Year2003         0.1988    0.1600   1.24 0.21415
## Year2004         0.1914    0.1639   1.17 0.24318
## Year2005         0.1527    0.1558   0.98 0.32725
## Year2006         0.2498    0.1515   1.65 0.09952 .
```

```

## Year2007          0.2069      0.1534      1.35  0.17753
## Year2008          0.2188      0.1502      1.46  0.14541
## Year2009          0.2632      0.1450      1.81  0.06982 .
## Year2010          0.1519      0.1496      1.01  0.31041
## Year2011          0.2177      0.1438      1.51  0.13028
## Year2012          0.2204      0.1425      1.55  0.12231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.0249
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.870   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      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 1.024 1          1.012
## Year              1.024 16          1.001

```



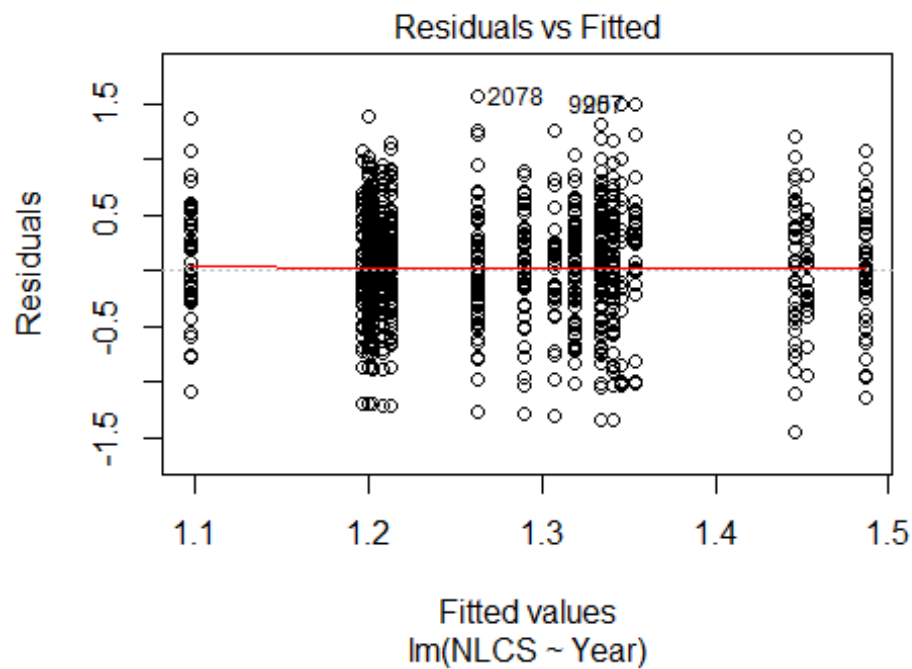
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5227 -0.4088 0.0118 0.4233 1.9101
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02673 0.13753 7.47 1.6e-13 ***
## LastAuthorFemale1 0.13798 0.05813 2.37 0.018 *
## Year1997 -0.09188 0.18143 -0.51 0.613
## Year1998 0.35112 0.16418 2.14 0.033 *
## Year1999 0.10528 0.18293 0.58 0.565
## Year2000 0.17481 0.16764 1.04 0.297
## Year2001 0.49596 0.18346 2.70 0.007 **
## Year2002 0.00501 0.16642 0.03 0.976
## Year2003 0.16348 0.16015 1.02 0.308
## Year2004 0.18255 0.16599 1.10 0.272
## Year2005 0.14536 0.15616 0.93 0.352
## Year2006 0.24733 0.15271 1.62 0.106
```

```

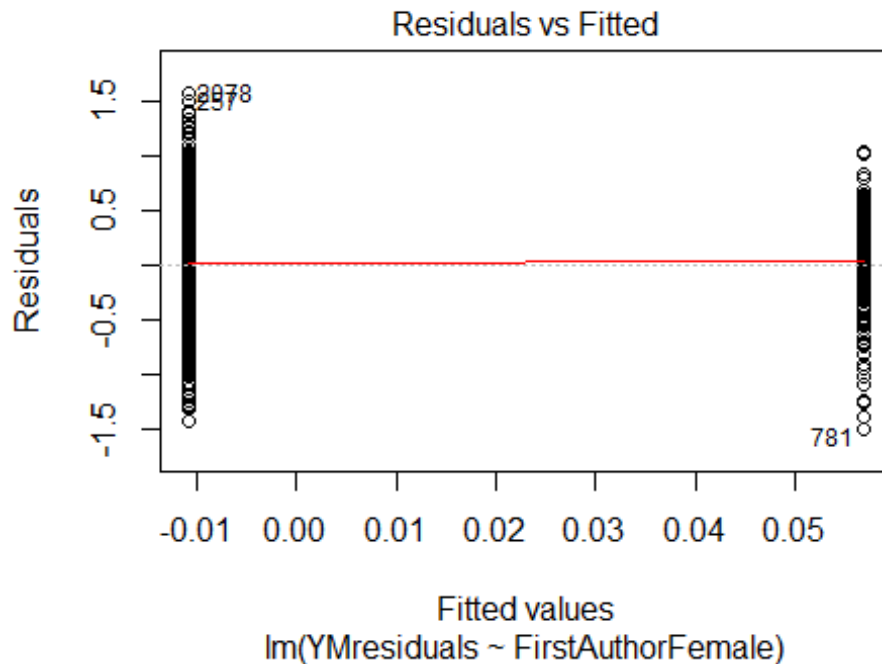
## Year2007      0.20489    0.15471    1.32    0.186
## Year2008      0.21450    0.15153    1.42    0.157
## Year2009      0.25550    0.14614    1.75    0.081 .
## Year2010      0.13967    0.15006    0.93    0.352
## Year2011      0.21553    0.14506    1.49    0.138
## Year2012      0.20442    0.14312    1.43    0.153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0212
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1129 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  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      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: 1243"
## [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
##  216  228  249  125  174  264  217  212  226  227  214  238  293  381  368
## 2011 2012
##  397  377
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   38   41   19   23   46   52   51   65   69   71   77   88  137  120
## 2011 2012

```

```
## 148 139
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 28 30 13 22 40 45 43 49 55 56 59 66 109 100
## 2011 2012
## 118 110
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

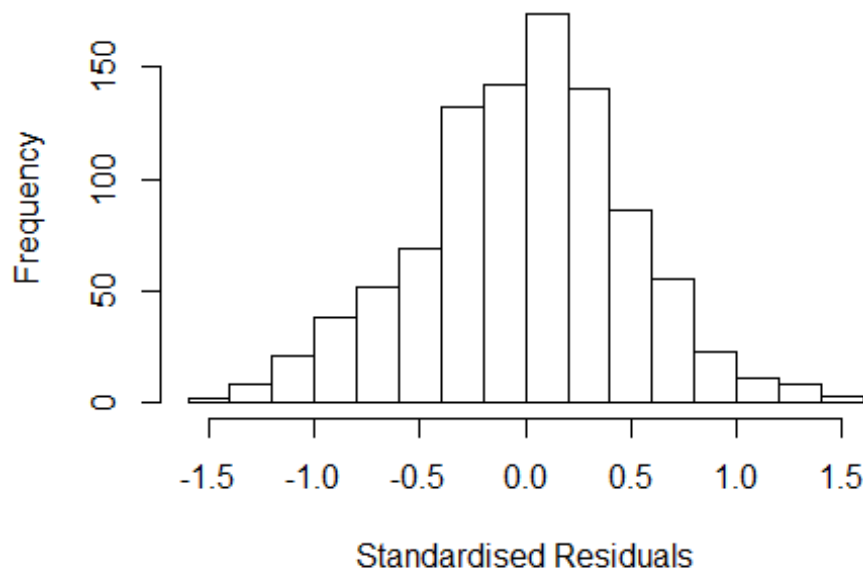


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.5, df = 1, p-value = 0.004
```

```
## [1] "Female first author team size 2018 geometric mean: 4.3056685594258"
## [1] "Male first author team size 2018 geometric mean: 2.97521557667979"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.05134407307113"
## [1] "Male last author team size 2018 geometric mean: 3.08258210943909"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 610, 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 1.170  1      1.082
## LastAuthorFemale  1.074  1      1.036
## UniqueAuthors    1.760  4      1.073
## Year             1.728 16      1.017
```

Residuals from first and last author and team size



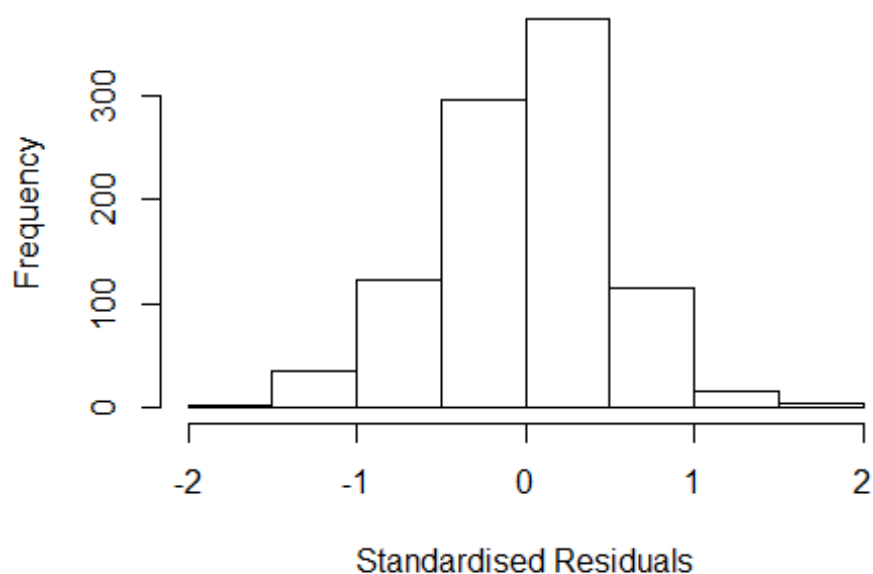
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5449 -0.3209  0.0185  0.3105  1.4988
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.2533    0.1134   11.05 < 2e-16 ***
## FirstAuthorFemale1  0.0474    0.0401    1.18  0.23752
## LastAuthorFemale1 -0.0211    0.0515   -0.41  0.68221
## UniqueAuthors2     0.2036    0.0616    3.30  0.00099 ***
## UniqueAuthors3     0.2763    0.0608    4.54  6.3e-06 ***
## UniqueAuthors4     0.2917    0.0699    4.17  3.3e-05 ***
## UniqueAuthors5     0.3783    0.0653    5.79  9.3e-09 ***
## Year1997         -0.0931    0.2003   -0.46  0.64232
## Year1998          0.0880    0.1497    0.59  0.55689
## Year1999         -0.1217    0.2785   -0.44  0.66234
```

```

## Year2000          -0.0257      0.1424   -0.18  0.85668
## Year2001           0.0625      0.1357    0.46  0.64500
## Year2002          -0.3698      0.1473   -2.51  0.01225 *
## Year2003          -0.0882      0.1315   -0.67  0.50268
## Year2004          -0.2078      0.1288   -1.61  0.10698
## Year2005          -0.3079      0.1257   -2.45  0.01449 *
## Year2006          -0.1642      0.1221   -1.34  0.17904
## Year2007          -0.1962      0.1277   -1.54  0.12489
## Year2008          -0.2722      0.1210   -2.25  0.02469 *
## Year2009          -0.1157      0.1143   -1.01  0.31172
## Year2010          -0.2765      0.1166   -2.37  0.01794 *
## Year2011          -0.3312      0.1156   -2.86  0.00428 **
## Year2012          -0.2607      0.1112   -2.34  0.01930 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.475
## Multiple R-squared:  0.0946, Adjusted R-squared:  0.0734
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 878 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.269  0.860  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           1.04e-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 1.077 1           1.038
## LastAuthorFemale  1.051 1           1.025
## Year              1.118 16           1.003

```

Residuals from first and last author



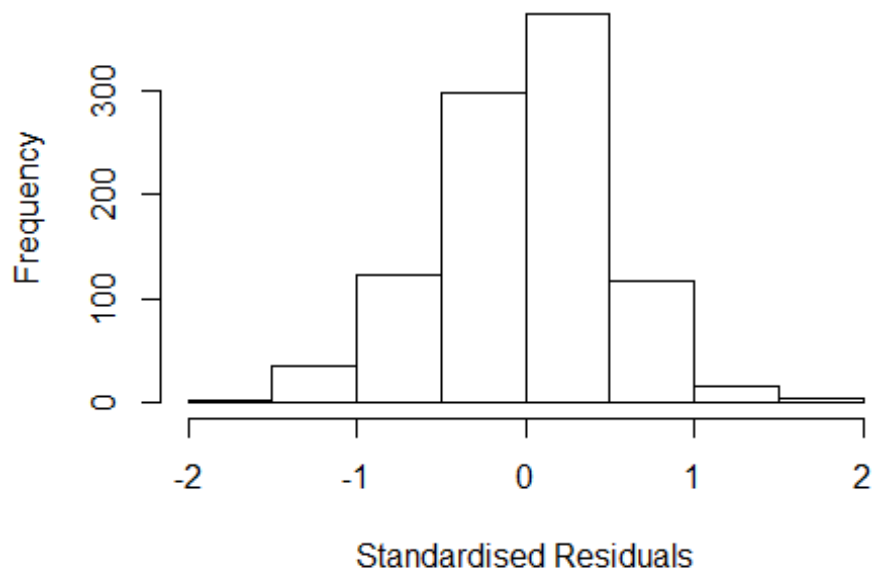
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5218 -0.3213  0.0288  0.3245  1.5955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.41000    0.09376   15.04  <2e-16 ***
## FirstAuthorFemale1 0.10754    0.03853    2.79   0.0054 **
## LastAuthorFemale1 -0.00467    0.05022   -0.09   0.9259
## Year1997          -0.07657    0.18580   -0.41   0.6804
## Year1998           0.11182    0.15013    0.74   0.4565
## Year1999          -0.13044    0.28491   -0.46   0.6472
## Year2000           0.01553    0.12931    0.12   0.9044
## Year2001           0.12161    0.12187    1.00   0.3186
## Year2002          -0.32271    0.13868   -2.33   0.0202 *
## Year2003          -0.09807    0.12266   -0.80   0.4242
## Year2004          -0.17752    0.12188   -1.46   0.1456
## Year2005          -0.27537    0.11720   -2.35   0.0190 *
```

```

## Year2006      -0.10546    0.11476   -0.92    0.3583
## Year2007      -0.11472    0.12087   -0.95    0.3428
## Year2008      -0.17403    0.11462   -1.52    0.1293
## Year2009      -0.02850    0.10468   -0.27    0.7855
## Year2010      -0.18065    0.10655   -1.70    0.0903 .
## Year2011      -0.24535    0.10606   -2.31    0.0209 *
## Year2012      -0.16948    0.10111   -1.68    0.0940 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.0502, Adjusted R-squared:  0.0321
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 885 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.865   0.952   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      1.04e-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 1.073 1      1.036
## Year              1.073 16      1.002

```

Residuals from first author



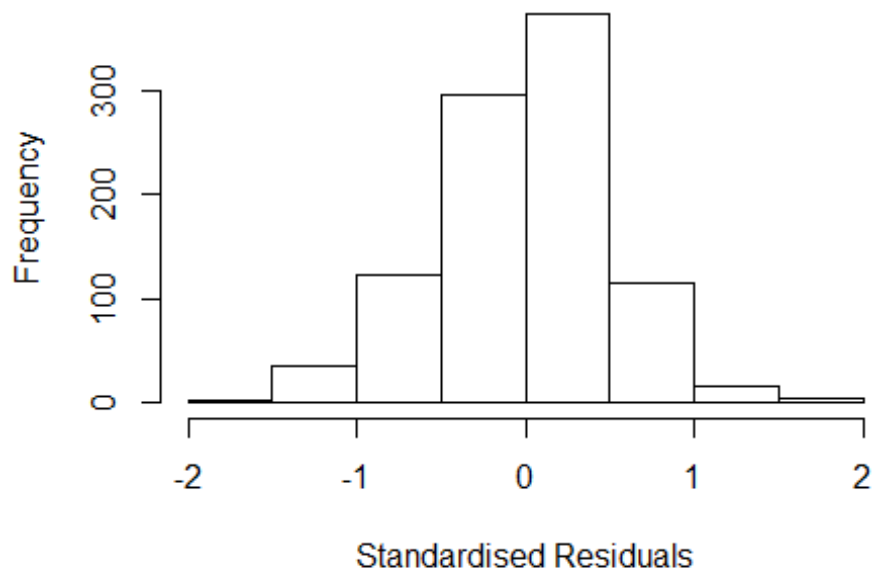
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5210 -0.3220 0.0284 0.3249 1.5959
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4098 0.0937 15.05 <2e-16 ***
## FirstAuthorFemale1 0.1072 0.0384 2.79 0.0054 **
## Year1997 -0.0765 0.1858 -0.41 0.6805
## Year1998 0.1113 0.1500 0.74 0.4585
## Year1999 -0.1313 0.2841 -0.46 0.6442
## Year2000 0.0151 0.1291 0.12 0.9072
## Year2001 0.1216 0.1219 1.00 0.3191
## Year2002 -0.3225 0.1386 -2.33 0.0202 *
## Year2003 -0.0983 0.1228 -0.80 0.4237
## Year2004 -0.1776 0.1220 -1.46 0.1457
## Year2005 -0.2756 0.1173 -2.35 0.0190 *
## Year2006 -0.1056 0.1147 -0.92 0.3576
```

```

## Year2007          -0.1150      0.1208   -0.95   0.3414
## Year2008          -0.1741      0.1146   -1.52   0.1292
## Year2009          -0.0288      0.1048   -0.27   0.7834
## Year2010          -0.1810      0.1066   -1.70   0.0898 .
## Year2011          -0.2456      0.1060   -2.32   0.0207 *
## Year2012          -0.1699      0.1012   -1.68   0.0935 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.0502, Adjusted R-squared:  0.0332
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 885 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.865   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      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 1.053 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
## -1.5424 -0.3199 0.0245 0.3221 1.5828
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.42080 0.09419 15.09 <2e-16 ***
## LastAuthorFemale1 0.00678 0.05040 0.13 0.893
## Year1997 -0.06965 0.18990 -0.37 0.714
## Year1998 0.12157 0.14727 0.83 0.409
## Year1999 -0.12951 0.29052 -0.45 0.656
## Year2000 0.02038 0.12920 0.16 0.875
## Year2001 0.11931 0.12230 0.98 0.330
## Year2002 -0.31526 0.13702 -2.30 0.022 *
## Year2003 -0.10460 0.12329 -0.85 0.396
## Year2004 -0.17561 0.12251 -1.43 0.152
## Year2005 -0.27445 0.11810 -2.32 0.020 *
## Year2006 -0.08864 0.11608 -0.76 0.445
```

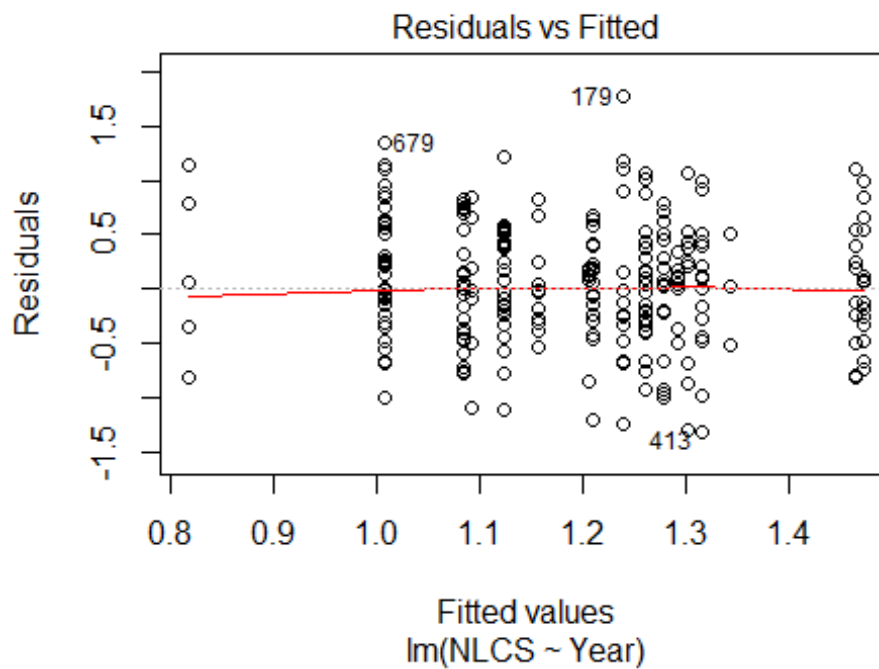


```

## Year2007          -0.11012      0.12129    -0.91      0.364
## Year2008          -0.16849      0.11494    -1.47      0.143
## Year2009          -0.01353      0.10461    -0.13      0.897
## Year2010          -0.16954      0.10669    -1.59      0.112
## Year2011          -0.22947      0.10635    -2.16      0.031 *
## Year2012          -0.16371      0.10124    -1.62      0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0271
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 891 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.260  0.862  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      1.04e-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: 964"
## [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
##   29   23   29   19   33   22   26   42   31   24   36   29   40   61   56
## 2011 2012
##   91   62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    9   10    7    3    6   16   15    9   11   19   15   24   24   18
## 2011 2012

```

```
## 48 35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 9 10 6 2 6 16 14 7 10 19 13 24 18 18
## 2011 2012
## 39 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 = 0.2
```



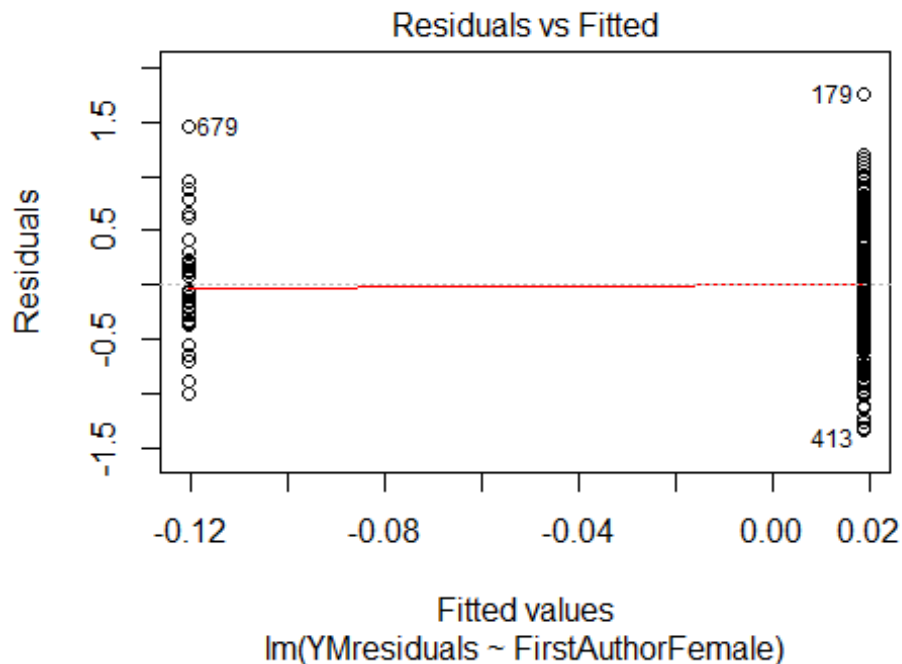
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.73, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 3.97030183847055"
## [1] "Male first author team size 2018 geometric mean: 2.66208399125877"

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

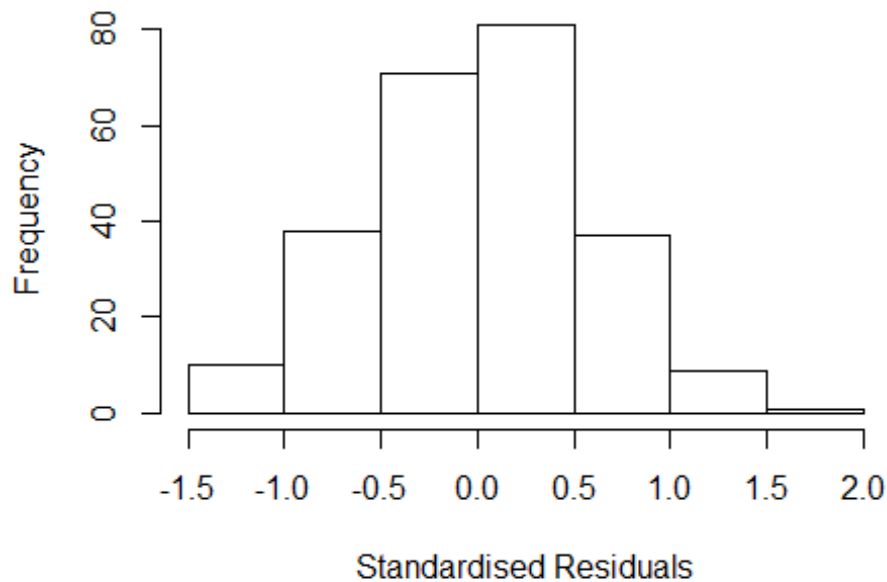
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.14376730109288"
## [1] "Male last author team size 2018 geometric mean: 2.7916773548895"

## 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 = 0.02
## 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 2.715  1      1.648
## LastAuthorFemale  2.098  1      1.448
## UniqueAuthors    4.273  4      1.199
## Year              8.231 16      1.068
```

Residuals from first and last author and team size



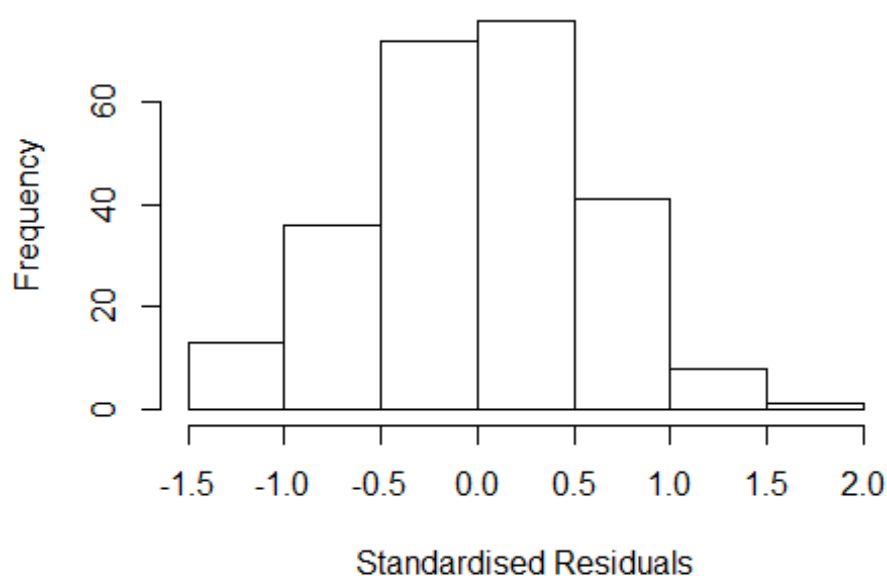
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3398 -0.3563 0.0201 0.3693 1.9527
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12335 0.27916 4.02 7.8e-05 ***
## FirstAuthorFemale1 -0.06993 0.12825 -0.55 0.59
## LastAuthorFemale1 -0.07575 0.12525 -0.60 0.55
## UniqueAuthors2 0.03661 0.11164 0.33 0.74
## UniqueAuthors3 0.12931 0.12052 1.07 0.28
## UniqueAuthors4 0.00223 0.15309 0.01 0.99
## UniqueAuthors5 0.21022 0.14558 1.44 0.15
## Year1997 0.21645 0.40111 0.54 0.59
## Year1998 0.32198 0.34089 0.94 0.35
## Year1999 0.06471 0.30787 0.21 0.83
```

```

## Year2000      0.14850    0.45612    0.33    0.75
## Year2001     -0.44902    0.49705   -0.90    0.37
## Year2002     -0.06501    0.38486   -0.17    0.87
## Year2003      0.33473    0.32086    1.04    0.30
## Year2004      0.08741    0.29722    0.29    0.77
## Year2005     -0.00243    0.31719   -0.01    0.99
## Year2006      0.15970    0.30222    0.53    0.60
## Year2007      0.11959    0.32329    0.37    0.71
## Year2008      0.13736    0.30138    0.46    0.65
## Year2009     -0.13276    0.29631   -0.45    0.65
## Year2010      0.03957    0.29896    0.13    0.89
## Year2011     -0.17844    0.29697   -0.60    0.55
## Year2012     -0.09630    0.30178   -0.32    0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0847, Adjusted R-squared:  -0.00523
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.209  0.864  0.957  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      4.05e-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 2.490 1      1.578
## LastAuthorFemale  1.961 1      1.400
## Year              2.273 16      1.026

```

Residuals from first and last author



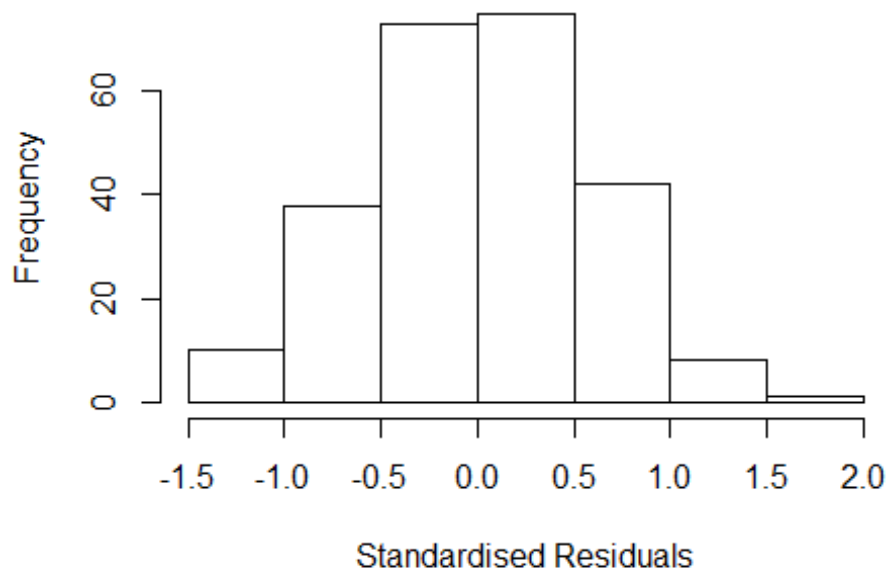
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3756 -0.3585 0.0406 0.3870 1.8592
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1384 0.2712 4.20 3.9e-05 ***
## FirstAuthorFemale1 -0.0935 0.1298 -0.72 0.47
## LastAuthorFemale1 -0.0751 0.1219 -0.62 0.54
## Year1997 0.2372 0.3957 0.60 0.55
## Year1998 0.3346 0.3360 1.00 0.32
## Year1999 0.0653 0.3066 0.21 0.83
## Year2000 0.1981 0.5084 0.39 0.70
## Year2001 -0.3827 0.5023 -0.76 0.45
## Year2002 0.0134 0.3585 0.04 0.97
## Year2003 0.3588 0.3171 1.13 0.26
## Year2004 0.1533 0.2854 0.54 0.59
## Year2005 0.0526 0.3140 0.17 0.87
```

```

## Year2006          0.1822      0.2999      0.61      0.54
## Year2007          0.1571      0.3263      0.48      0.63
## Year2008          0.1510      0.2975      0.51      0.61
## Year2009         -0.0910      0.2966     -0.31      0.76
## Year2010          0.1102      0.2921      0.38      0.71
## Year2011         -0.1304      0.2958     -0.44      0.66
## Year2012         -0.0246      0.2933     -0.08      0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.0727, Adjusted R-squared:  -0.000556
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.859  0.954  0.908  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      4.05e-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 1.915 1      1.384
## Year              1.915 16      1.021

```

Residuals from first author



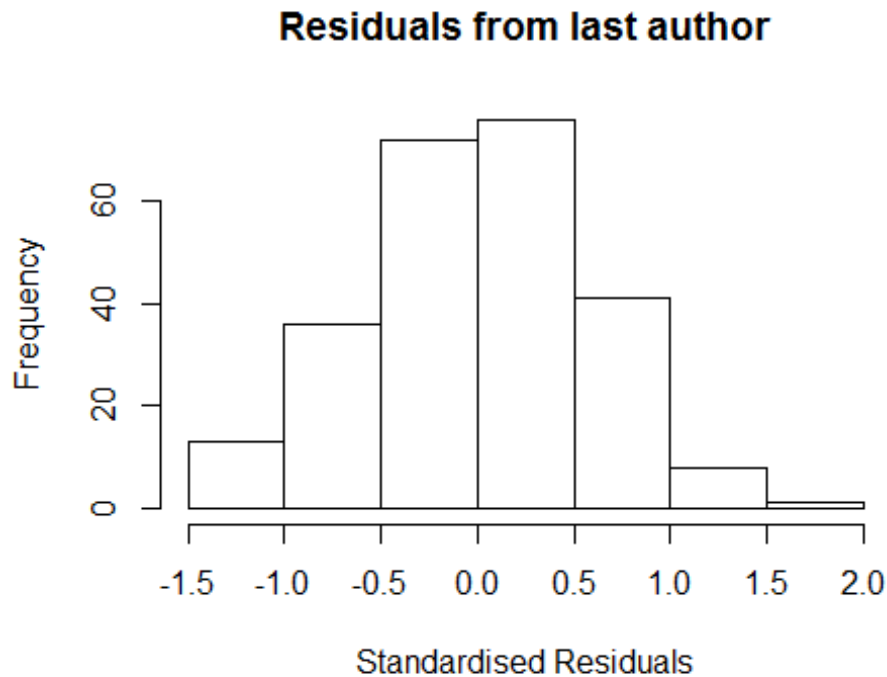
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3745 -0.3620 0.0244 0.3903 1.8597
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1330 0.2662 4.26 3e-05 ***
## FirstAuthorFemale1 -0.1234 0.1153 -1.07 0.29
## Year1997 0.2416 0.3919 0.62 0.54
## Year1998 0.3398 0.3342 1.02 0.31
## Year1999 0.0704 0.3023 0.23 0.82
## Year2000 0.2035 0.5041 0.40 0.69
## Year2001 -0.3752 0.4814 -0.78 0.44
## Year2002 0.0184 0.3530 0.05 0.96
## Year2003 0.3564 0.3117 1.14 0.25
## Year2004 0.1467 0.2829 0.52 0.60
## Year2005 0.0353 0.3074 0.11 0.91
## Year2006 0.1830 0.2957 0.62 0.54
```



```

## Year2007      0.1515      0.3211      0.47      0.64
## Year2008      0.1470      0.2913      0.50      0.61
## Year2009     -0.0900      0.2920     -0.31      0.76
## Year2010      0.1096      0.2871      0.38      0.70
## Year2011     -0.1344      0.2895     -0.46      0.64
## Year2012     -0.0192      0.2888     -0.07      0.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.575
## Multiple R-squared:  0.071, Adjusted R-squared:  0.00203
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 228 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.275  0.864  0.953  0.908  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      4.05e-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 1.447 1      1.203
## Year      1.447 16      1.012

```



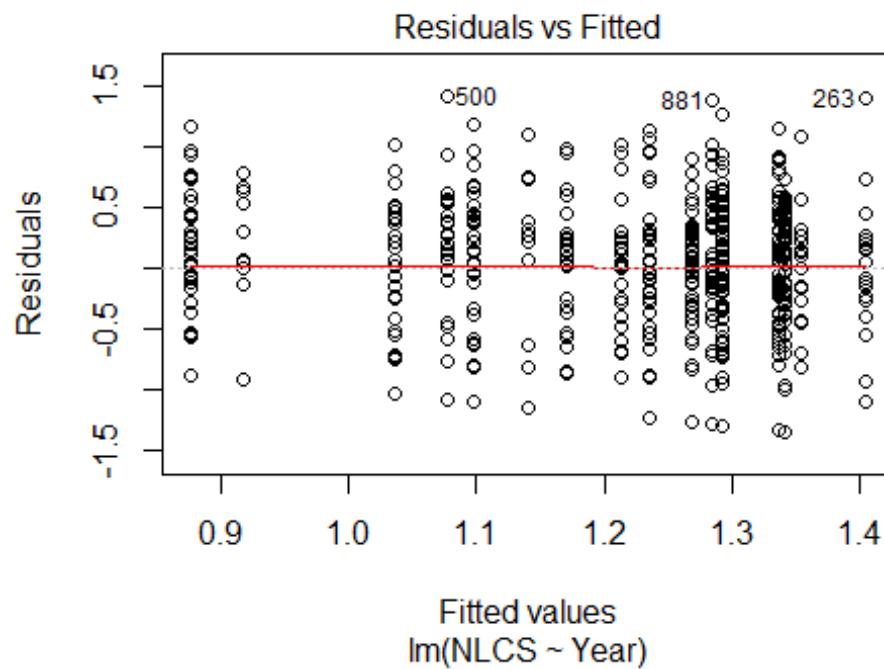
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3768 -0.3753 0.0439 0.3833 1.8768
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13232 0.26575 4.26 3e-05 ***
## LastAuthorFemale1 -0.11245 0.10822 -1.04 0.30
## Year1997 0.24449 0.39237 0.62 0.53
## Year1998 0.32794 0.32990 0.99 0.32
## Year1999 0.07172 0.30171 0.24 0.81
## Year2000 0.20418 0.50632 0.40 0.69
## Year2001 -0.37950 0.50687 -0.75 0.45
## Year2002 0.00185 0.35496 0.01 1.00
## Year2003 0.35466 0.31149 1.14 0.26
## Year2004 0.14362 0.27802 0.52 0.61
## Year2005 0.03369 0.30393 0.11 0.91
## Year2006 0.18578 0.29472 0.63 0.53
```

```

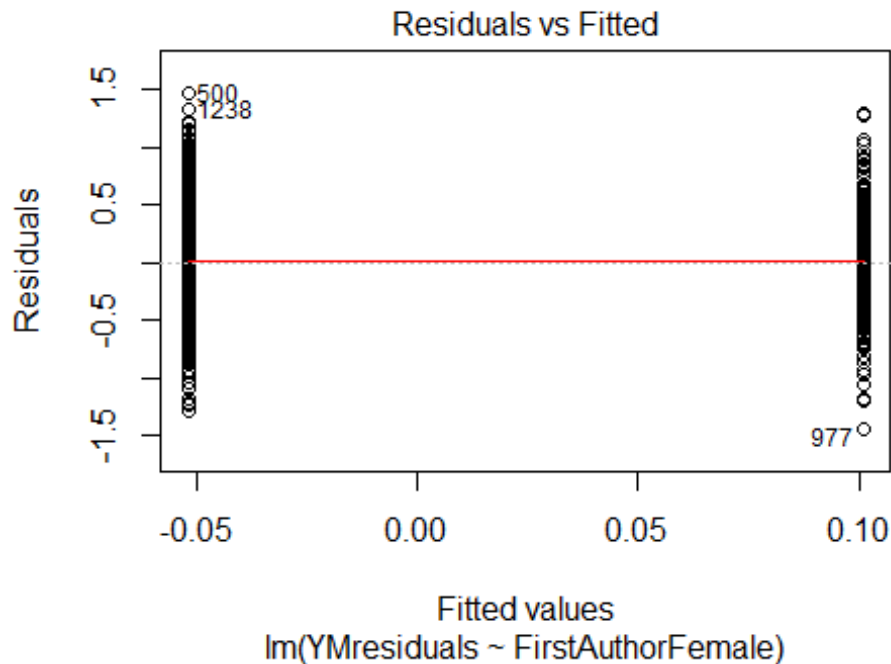
## Year2007      0.13739      0.31640      0.43      0.66
## Year2008      0.13828      0.29026      0.48      0.63
## Year2009     -0.08487      0.29024     -0.29      0.77
## Year2010      0.10963      0.28648      0.38      0.70
## Year2011     -0.12509      0.29054     -0.43      0.67
## Year2012     -0.02192      0.28794     -0.08      0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0703, Adjusted R-squared:  0.00124
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.254  0.856  0.952  0.906  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      4.05e-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 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
##   54   45   51   51   60   66   55   48   44   38   61   72   76   69   97
## 2011 2012
##  107  104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   16   13   24   22   17   18   29   30   24   33   46   47   36   64
## 2011 2012

```

```
## 71 70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 12 10 20 21 14 16 25 28 23 31 46 40 31 56
## 2011 2012
## 64 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 = 18, df = 16, p-value = 0.3
```

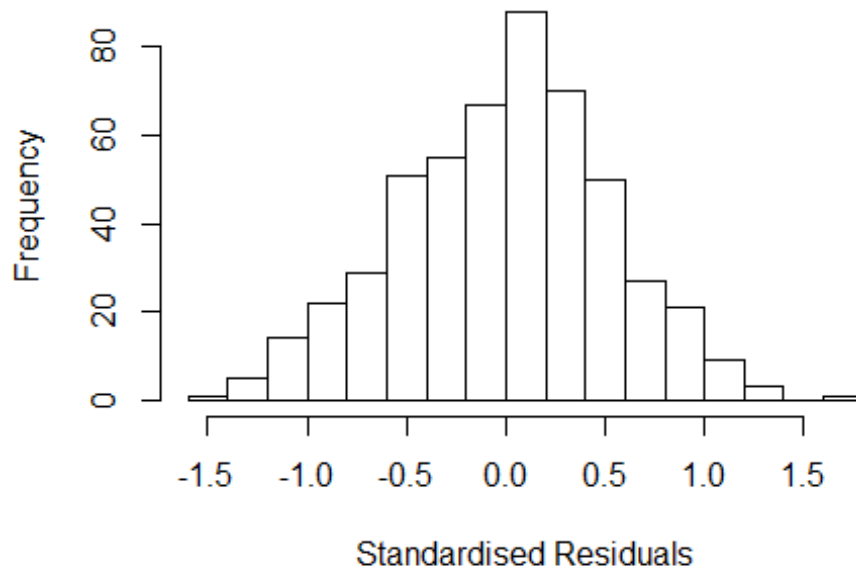


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.32244314109422"
## [1] "Male first author team size 2018 geometric mean: 2.96950017324914"
##
## 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] "Female last author team size 2018 geometric mean: 2.94153200905655"
## [1] "Male last author team size 2018 geometric mean: 3.15049980785066"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 740, 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 1.464 1      1.210
## LastAuthorFemale  1.365 1      1.168
## UniqueAuthors     1.997 4      1.090
## Year              2.360 16     1.027
```

Residuals from first and last author and team size



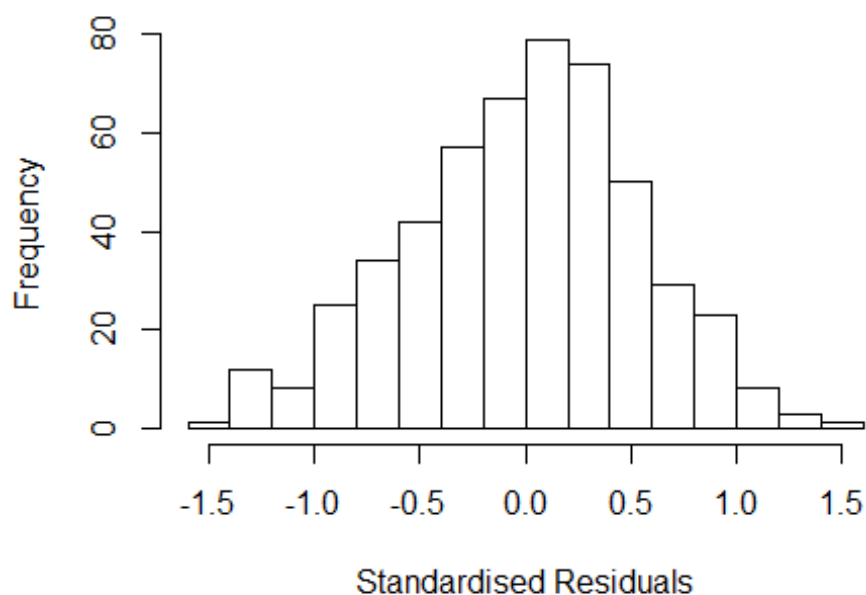
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4127 -0.3822 0.0221 0.3688 1.6431
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1362 0.1414 8.03 7.1e-15 ***
## FirstAuthorFemale1 0.1056 0.0584 1.81 0.0711 .
## LastAuthorFemale1 0.0973 0.0595 1.63 0.1028
## UniqueAuthors2 0.0804 0.0717 1.12 0.2629
## UniqueAuthors3 0.2907 0.0731 3.97 8.1e-05 ***
## UniqueAuthors4 0.2582 0.0792 3.26 0.0012 **
## UniqueAuthors5 0.3171 0.0970 3.27 0.0012 **
## Year1997 -0.1463 0.2456 -0.60 0.5518
## Year1998 -0.2318 0.2212 -1.05 0.2953
## Year1999 -0.0456 0.1830 -0.25 0.8035
```

```

## Year2000          0.0968      0.1721      0.56      0.5743
## Year2001          0.1547      0.2044      0.76      0.4496
## Year2002          0.0463      0.2926      0.16      0.8742
## Year2003         -0.3047      0.1947     -1.56      0.1183
## Year2004         -0.2923      0.1908     -1.53      0.1262
## Year2005         -0.0836      0.1805     -0.46      0.6434
## Year2006         -0.1755      0.1918     -0.92      0.3605
## Year2007         -0.4303      0.1689     -2.55      0.0111 *
## Year2008         -0.0838      0.1616     -0.52      0.6043
## Year2009         -0.0143      0.1951     -0.07      0.9418
## Year2010         -0.0453      0.1651     -0.27      0.7840
## Year2011         -0.0129      0.1641     -0.08      0.9372
## Year2012         -0.1058      0.1607     -0.66      0.5107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.145, Adjusted R-squared:  0.106
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 470 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.314  0.866  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      1.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.410 1      1.188
## LastAuthorFemale  1.406 1      1.186
## Year              1.324 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
## -1.4682 -0.3727 0.0397 0.3681 1.5637
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20549 0.14404 8.37 6e-16 ***
## FirstAuthorFemale1 0.14319 0.05959 2.40 0.017 *
## LastAuthorFemale1 0.09274 0.06231 1.49 0.137
## Year1997 -0.12041 0.25830 -0.47 0.641
## Year1998 -0.27974 0.22786 -1.23 0.220
## Year1999 -0.05434 0.18624 -0.29 0.771
## Year2000 0.15772 0.17221 0.92 0.360
## Year2001 0.12667 0.19120 0.66 0.508
## Year2002 0.05144 0.30311 0.17 0.865
## Year2003 -0.28118 0.19119 -1.47 0.142
## Year2004 -0.28221 0.19452 -1.45 0.147
## Year2005 -0.05318 0.18287 -0.29 0.771
```

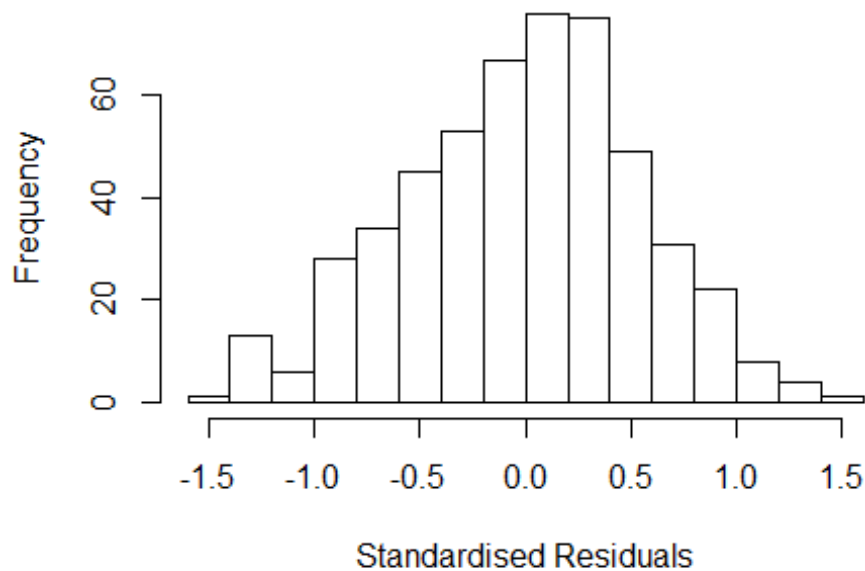


```

## Year2006          -0.12615      0.18929      -0.67      0.505
## Year2007          -0.39362      0.16935      -2.32      0.021 *
## Year2008          -0.01800      0.16022      -0.11      0.911
## Year2009           0.03603      0.19524       0.18      0.854
## Year2010           0.02679      0.16253       0.16      0.869
## Year2011           0.09551      0.16201       0.59      0.556
## Year2012          -0.00117      0.15993      -0.01      0.994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0686
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 471 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.386  0.877   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      1.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##           "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.164 1          1.079
## Year              1.164 16          1.005

```

Residuals from first author



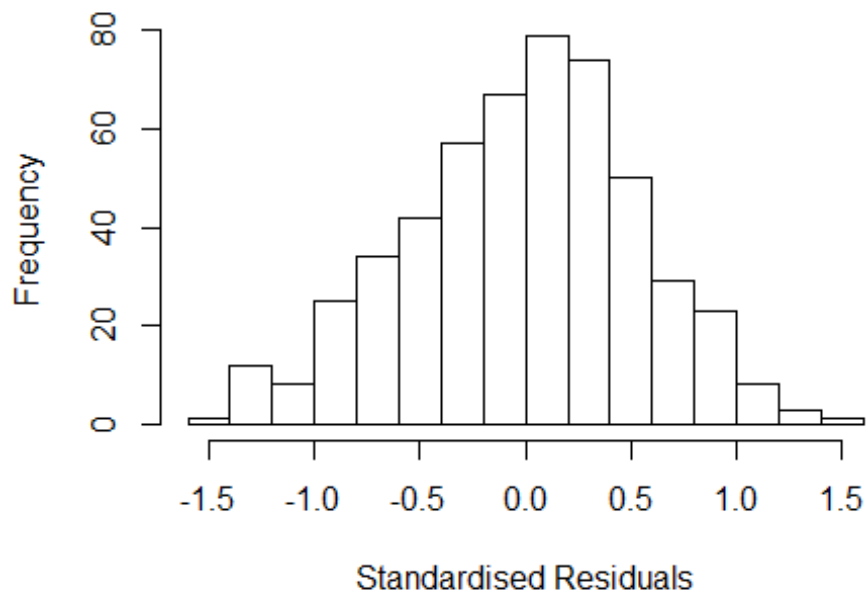
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.439 -0.394 0.042 0.355 1.549
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21854 0.14534 8.38 5.3e-16 ***
## FirstAuthorFemale1 0.18222 0.05381 3.39 0.00076 ***
## Year1997 -0.13326 0.26339 -0.51 0.61313
## Year1998 -0.28585 0.22803 -1.25 0.21060
## Year1999 -0.05589 0.18828 -0.30 0.76671
## Year2000 0.15078 0.17429 0.87 0.38739
## Year2001 0.12460 0.19209 0.65 0.51684
## Year2002 0.03976 0.29993 0.13 0.89460
## Year2003 -0.28662 0.19031 -1.51 0.13268
## Year2004 -0.28049 0.19584 -1.43 0.15271
## Year2005 -0.05768 0.18274 -0.32 0.75240
## Year2006 -0.11860 0.19016 -0.62 0.53312
```

```

## Year2007          -0.40484    0.17083   -2.37  0.01817 *
## Year2008          -0.02159    0.16166   -0.13  0.89381
## Year2009           0.03116    0.19576    0.16  0.87360
## Year2010           0.03803    0.16250    0.23  0.81504
## Year2011           0.09818    0.16300    0.60  0.54724
## Year2012          -0.00729    0.16107   -0.05  0.96392
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.097, Adjusted R-squared:  0.066
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.410  0.879  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      1.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.149 1          1.072
## Year              1.149 16          1.004

```

Residuals from last author



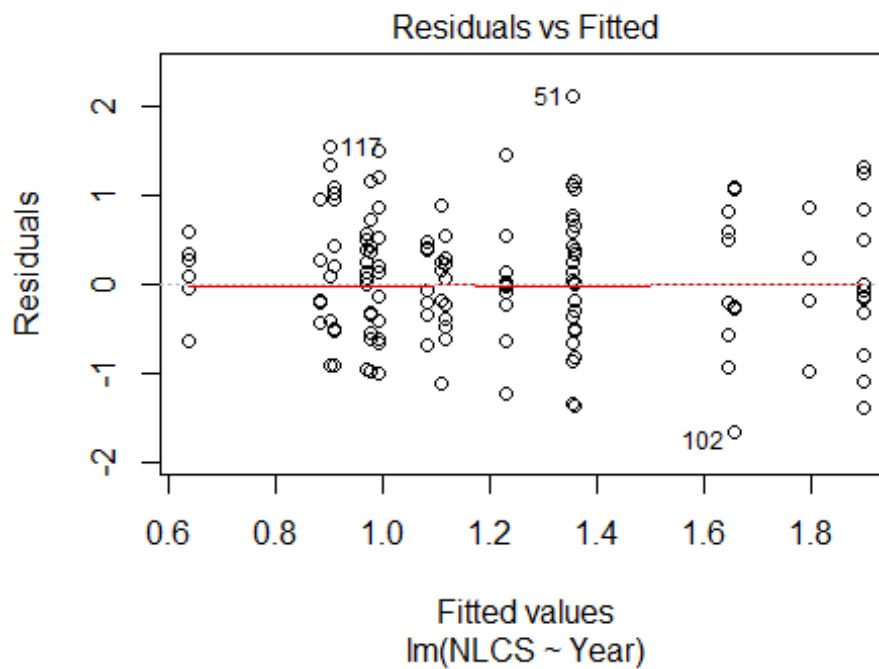
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4191 -0.3773 0.0294 0.3877 1.5102
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20991 0.14744 8.21 2e-15 ***
## LastAuthorFemale1 0.15982 0.05594 2.86 0.0045 **
## Year1997 -0.11103 0.25743 -0.43 0.6664
## Year1998 -0.25050 0.23548 -1.06 0.2879
## Year1999 -0.04731 0.19101 -0.25 0.8045
## Year2000 0.17739 0.17635 1.01 0.3150
## Year2001 0.14189 0.19400 0.73 0.4649
## Year2002 0.05756 0.30283 0.19 0.8493
## Year2003 -0.25516 0.19278 -1.32 0.1862
## Year2004 -0.23316 0.19726 -1.18 0.2378
## Year2005 -0.04306 0.18672 -0.23 0.8177
## Year2006 -0.12217 0.19174 -0.64 0.5243
```

```

## Year2007          -0.35956      0.17248      -2.08      0.0376 *
## Year2008           0.00974      0.16381       0.06      0.9526
## Year2009           0.06337      0.19613       0.32      0.7467
## Year2010           0.04940      0.16654       0.30      0.7669
## Year2011           0.13202      0.16432       0.80      0.4221
## Year2012           0.04394      0.16216       0.27      0.7865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0913, Adjusted R-squared:  0.06
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 469 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.418  0.869  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      1.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 513"
## [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
##   14   15   20    9    8   18   12   11    7   10    9   12   17   19   19
## 2011 2012
##   16   16
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12   13   18    6    7    5    8    4    5    7    7    9   13    9   10
## 2011 2012

```

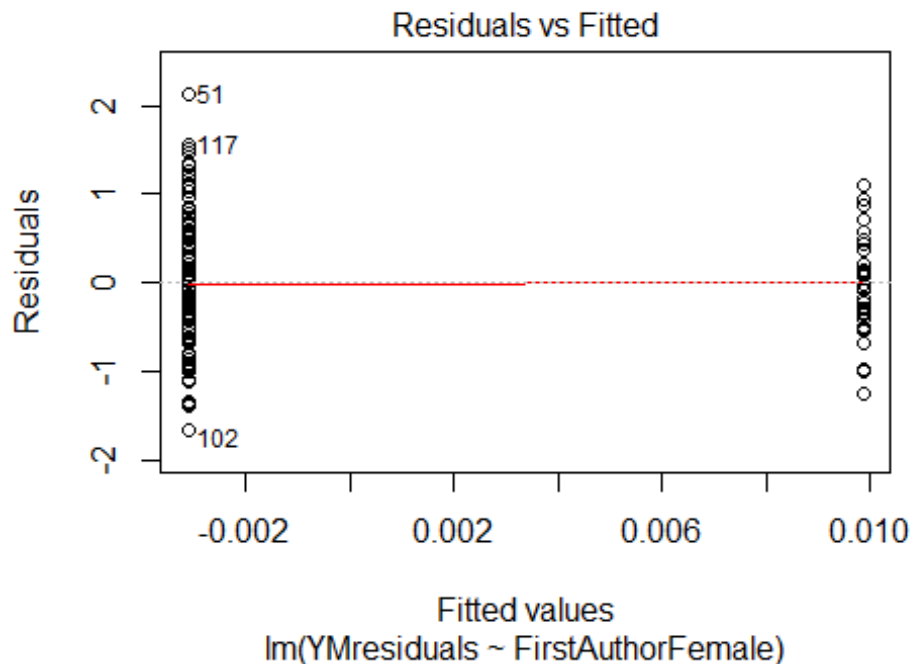
```
## 11 10
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 11 13 4 7 3 6 4 3 6 7 7 12 8 8
## 2011 2012
## 11 6
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



```
##
## 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: 1.5874010519682"
## [1] "Male first author team size 2018 geometric mean: 1.64582191221602"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

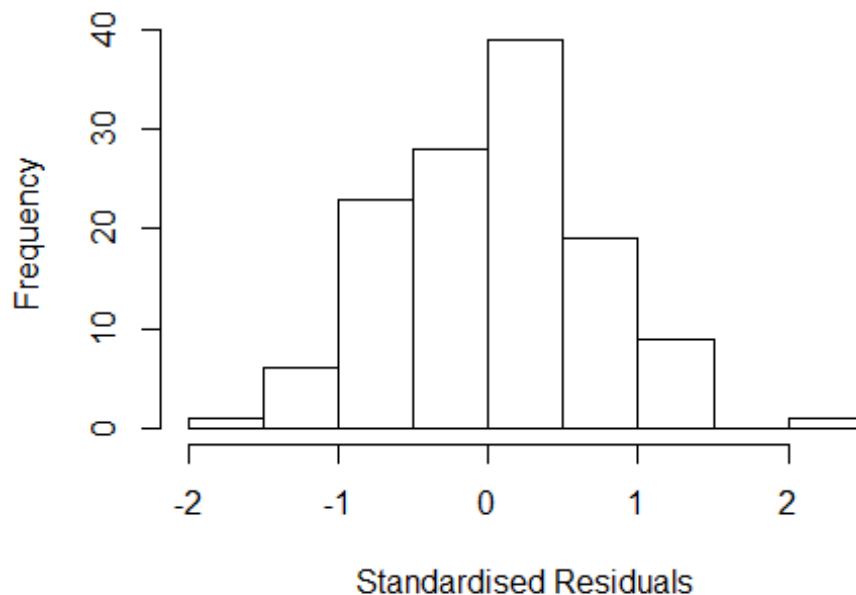
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2"
## [1] "Male last author team size 2018 geometric mean: 1.50596585461492"

## 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 = 0.6
## 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  3.419  1      1.849
## LastAuthorFemale  3.359  1      1.833
## UniqueAuthors    29.948  4      1.529
## Year              56.959 16      1.135
```

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
## -1.5306 -0.4419 0.0604 0.4543 2.1436
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4513 0.4171 3.48 0.00074 ***
## FirstAuthorFemale1 -0.0811 0.1869 -0.43 0.66533
## LastAuthorFemale1 -0.0573 0.2156 -0.27 0.79093
## UniqueAuthors2 0.5046 0.2128 2.37 0.01958 *
## UniqueAuthors3 0.4181 0.2020 2.07 0.04096 *
## UniqueAuthors4 0.4508 0.2666 1.69 0.09382 .
## UniqueAuthors5 0.7600 0.3480 2.18 0.03125 *
## Year1997 -0.3388 0.4845 -0.70 0.48598
## Year1998 -0.6226 0.4791 -1.30 0.19667
## Year1999 -0.7694 0.5527 -1.39 0.16686
```

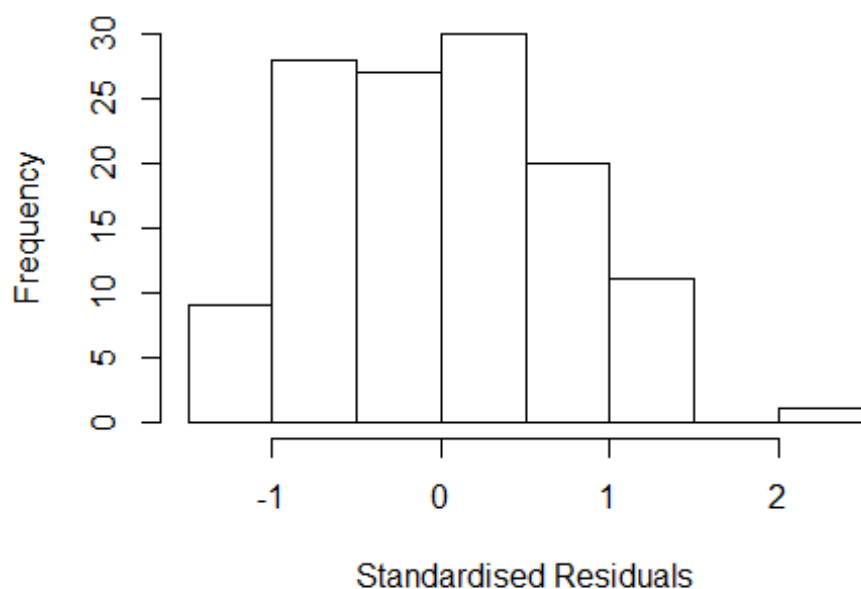


```

## Year2000          -0.1656      0.4533   -0.37   0.71557
## Year2001          -0.9377      0.4621   -2.03   0.04504 *
## Year2002          -0.5686      0.5205   -1.09   0.27725
## Year2003           0.1388      0.5408    0.26   0.79790
## Year2004          -0.6881      0.4254   -1.62   0.10884
## Year2005          -1.2519      0.4530   -2.76   0.00678 **
## Year2006          -0.5015      0.4286   -1.17   0.24466
## Year2007          -0.7076      0.4285   -1.65   0.10169
## Year2008          -0.7992      0.4293   -1.86   0.06553 .
## Year2009          -0.8684      0.3951   -2.20   0.03022 *
## Year2010          -0.5743      0.4544   -1.26   0.20914
## Year2011          -0.6635      0.4436   -1.50   0.13774
## Year2012          -0.3826      0.4754   -0.80   0.42280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.728
## Multiple R-squared:  0.242, Adjusted R-squared:  0.0802
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.366  0.873  0.956  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           7.94e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.606 1           1.614
## LastAuthorFemale  2.294 1           1.515
## Year              2.355 16           1.027

```

Residuals from first and last author



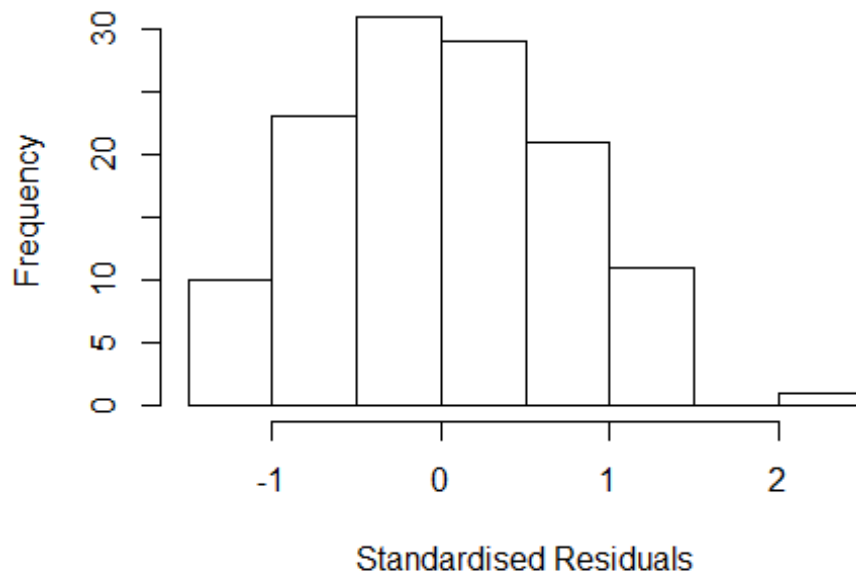
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4169 -0.5641 -0.0201 0.5058 2.2113
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.9299 0.3303 5.84 5.6e-08 ***
## FirstAuthorFemale1 0.0123 0.1929 0.06 0.94933
## LastAuthorFemale1 -0.1522 0.2303 -0.66 0.51013
## Year1997 -0.5415 0.4131 -1.31 0.19269
## Year1998 -0.6642 0.4416 -1.50 0.13549
## Year1999 -0.9006 0.4347 -2.07 0.04069 *
## Year2000 -0.2553 0.4072 -0.63 0.53207
## Year2001 -0.9688 0.5323 -1.82 0.07157 .
## Year2002 -0.7486 0.5679 -1.32 0.19025
## Year2003 -0.0658 0.4801 -0.14 0.89123
## Year2004 -0.7488 0.3477 -2.15 0.03352 *
## Year2005 -1.3659 0.3732 -3.66 0.00039 ***
```

```

## Year2006          -0.7836      0.3609    -2.17  0.03213 *
## Year2007          -0.8462      0.3752    -2.26  0.02617 *
## Year2008          -1.0441      0.3926    -2.66  0.00903 **
## Year2009          -0.9258      0.3759    -2.46  0.01537 *
## Year2010          -0.7761      0.4905    -1.58  0.11650
## Year2011          -0.9358      0.3959    -2.36  0.01989 *
## Year2012          -0.6472      0.4463    -1.45  0.14994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.829
## Multiple R-squared:  0.171, Adjusted R-squared:  0.0312
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 117 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.457  0.903  0.958  0.930  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      7.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.629 1      1.276
## Year              1.629 16      1.015

```

Residuals from first author

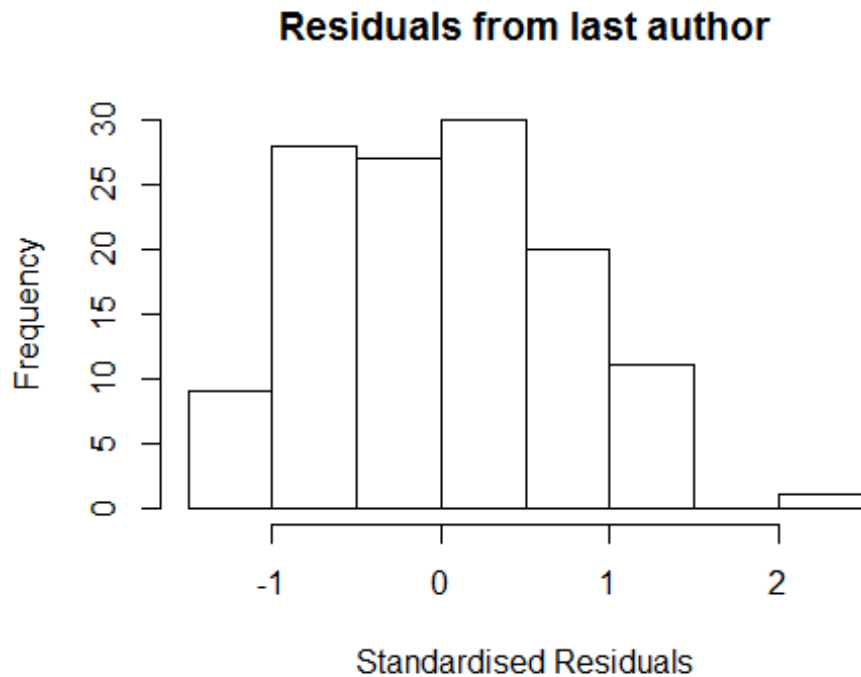


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39604 -0.55099 -0.00829 0.50870 2.20635
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.9065 0.3382 5.64 1.4e-07 ***
## FirstAuthorFemale1 -0.0631 0.1522 -0.41 0.67941
## Year1997 -0.5104 0.4156 -1.23 0.22205
## Year1998 -0.6358 0.4508 -1.41 0.16131
## Year1999 -0.8567 0.4306 -1.99 0.04917 *
## Year2000 -0.2569 0.4182 -0.61 0.54032
## Year2001 -0.9459 0.5368 -1.76 0.08087 .
## Year2002 -0.7196 0.5406 -1.33 0.18594
## Year2003 -0.0737 0.4884 -0.15 0.88031
## Year2004 -0.7253 0.3552 -2.04 0.04360 *
## Year2005 -1.3541 0.3757 -3.60 0.00047 ***
## Year2006 -0.7936 0.3703 -2.14 0.03435 *
```

```

## Year2007          -0.8340      0.3817   -2.19  0.03105 *
## Year2008          -1.0364      0.3960   -2.62  0.01013 *
## Year2009          -0.9249      0.3941   -2.35  0.02077 *
## Year2010          -0.7569      0.4887   -1.55  0.12437
## Year2011          -0.9142      0.4031   -2.27  0.02534 *
## Year2012          -0.6637      0.4578   -1.45  0.15005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.836
## Multiple R-squared:  0.167, Adjusted R-squared:  0.0354
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 118 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.466  0.898  0.961  0.931  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      7.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.45 1          1.204
## Year              1.45 16          1.012

```



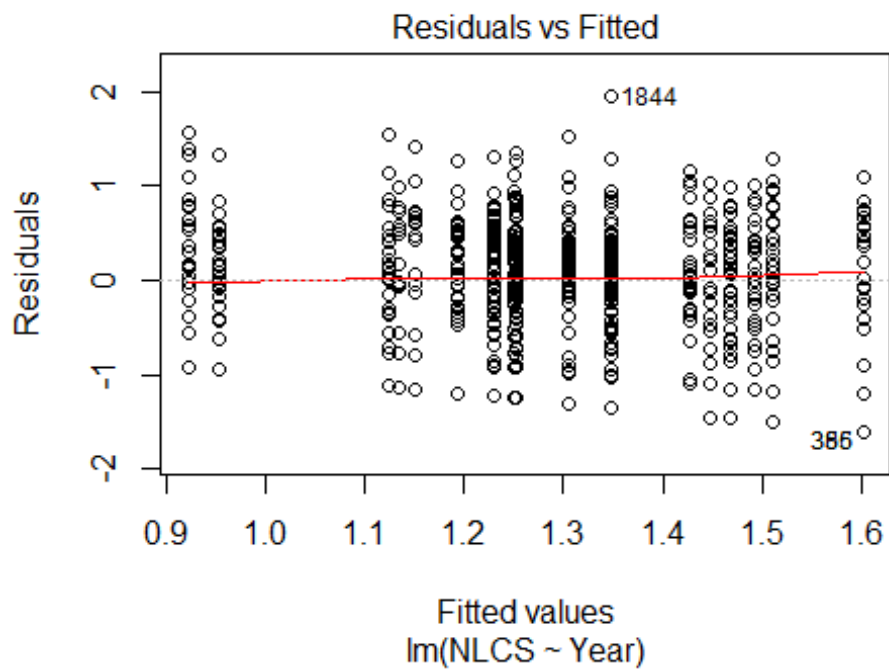
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4170 -0.5642 -0.0223 0.5001 2.2096
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.9300 0.3305 5.84 5.6e-08 ***
## LastAuthorFemale1 -0.1451 0.1849 -0.78 0.43432
## Year1997 -0.5404 0.4130 -1.31 0.19355
## Year1998 -0.6626 0.4410 -1.50 0.13587
## Year1999 -0.8974 0.4311 -2.08 0.03972 *
## Year2000 -0.2566 0.4055 -0.63 0.52829
## Year2001 -0.9689 0.5327 -1.82 0.07169 .
## Year2002 -0.7462 0.5577 -1.34 0.18370
## Year2003 -0.0638 0.4840 -0.13 0.89543
## Year2004 -0.7489 0.3479 -2.15 0.03359 *
## Year2005 -1.3649 0.3731 -3.66 0.00039 ***
## Year2006 -0.7814 0.3624 -2.16 0.03329 *
```

```

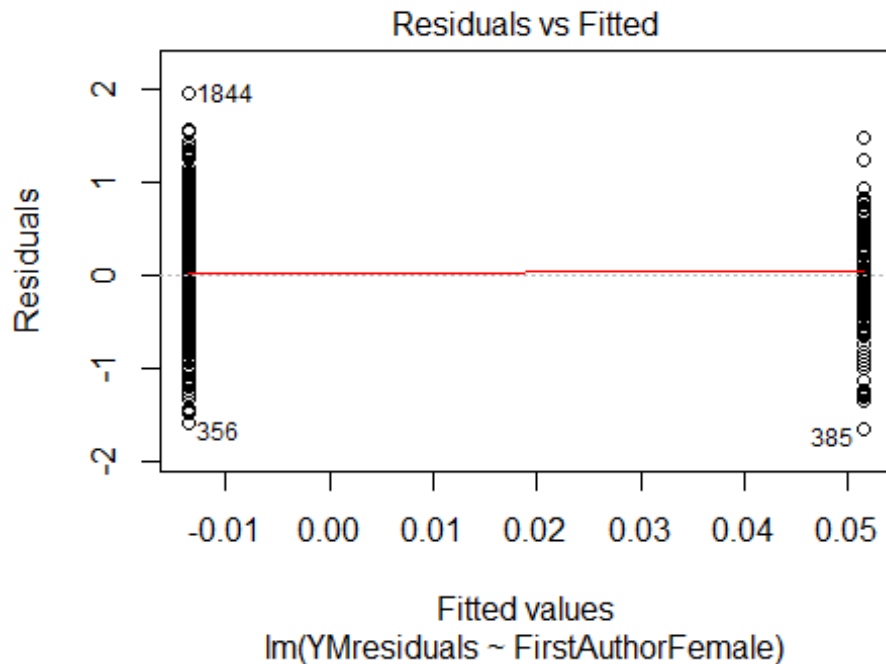
## Year2007          -0.8455      0.3761   -2.25  0.02660 *
## Year2008          -1.0411      0.3904   -2.67  0.00884 **
## Year2009          -0.9237      0.3753   -2.46  0.01543 *
## Year2010          -0.7702      0.4702   -1.64  0.10434
## Year2011          -0.9325      0.3943   -2.37  0.01981 *
## Year2012          -0.6445      0.4466   -1.44  0.15182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.83
## Multiple R-squared:  0.171, Adjusted R-squared:  0.0401
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 117 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.458  0.904   0.958   0.930   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      7.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 126"
## [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
##   34   56   60   68   75   76   66   68   95   74   66   74   86  112  126
## 2011 2012
##  139  164
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   21   21   29   29   24   36   33   37   33   42   38   37   55   61
## 2011 2012

```

```
## 64 78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 20 17 23 25 23 32 30 32 26 34 30 34 46 56
## 2011 2012
## 50 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 = 18, df = 16, p-value = 0.3
```

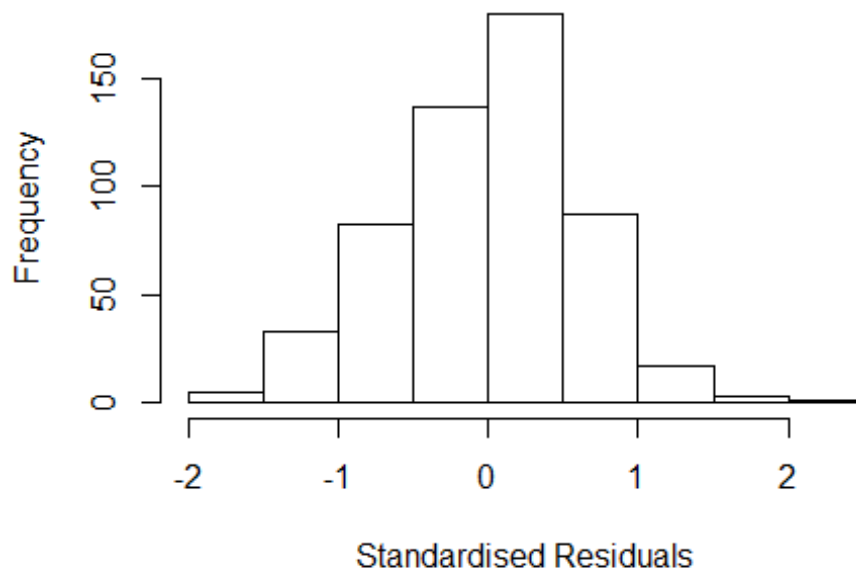


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```

```
## [1] "Female first author team size 2018 geometric mean: 3.33915125136442"
## [1] "Male first author team size 2018 geometric mean: 2.57854770686154"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.76575735439952"
## [1] "Male last author team size 2018 geometric mean: 2.71516201666152"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, 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 1.555 1      1.247
## LastAuthorFemale  1.276 1      1.130
## UniqueAuthors     1.699 4      1.069
## Year              2.067 16      1.023
```

Residuals from first and last author and team size



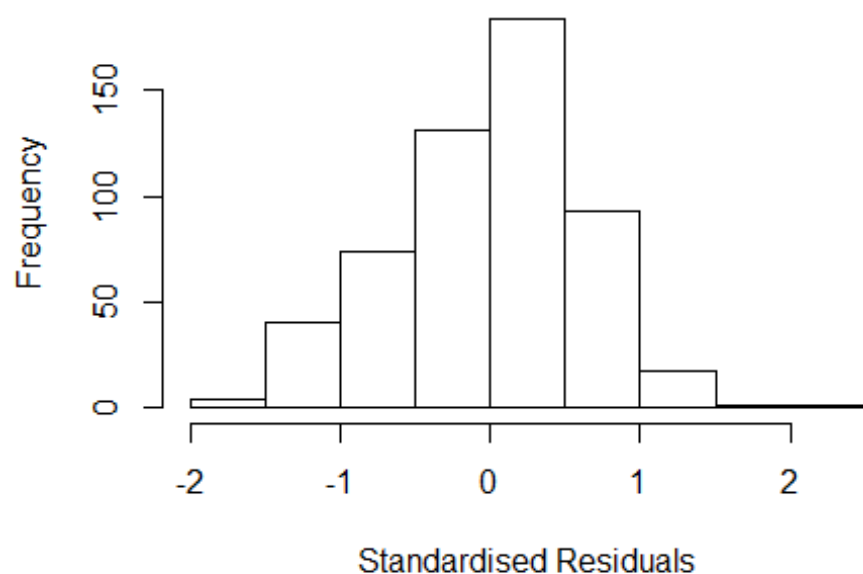
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8523 -0.4096 0.0403 0.3988 2.0257
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1156 0.1827 6.11 2.0e-09 ***
## FirstAuthorFemale1 0.1059 0.0735 1.44 0.15001
## LastAuthorFemale1 0.0593 0.0731 0.81 0.41731
## UniqueAuthors2 0.2531 0.0693 3.65 0.00028 ***
## UniqueAuthors3 0.2591 0.0783 3.31 0.00100 **
## UniqueAuthors4 0.1906 0.1118 1.70 0.08892 .
## UniqueAuthors5 0.5558 0.1281 4.34 1.7e-05 ***
## Year1997 -0.1465 0.2907 -0.50 0.61464
## Year1998 0.2894 0.2284 1.27 0.20578
## Year1999 0.2174 0.2156 1.01 0.31376
```

```

## Year2000          0.2101      0.2455      0.86  0.39263
## Year2001          0.3777      0.2511      1.50  0.13317
## Year2002         -0.4016      0.2367     -1.70  0.09027 .
## Year2003         -0.2216      0.2344     -0.95  0.34507
## Year2004          0.0198      0.2085      0.10  0.92434
## Year2005         -0.4236      0.2139     -1.98  0.04813 *
## Year2006         -0.0215      0.1996     -0.11  0.91409
## Year2007         -0.1550      0.2174     -0.71  0.47616
## Year2008          0.2439      0.2032      1.20  0.23064
## Year2009          0.0638      0.2031      0.31  0.75349
## Year2010         -0.1228      0.1941     -0.63  0.52741
## Year2011         -0.0355      0.1974     -0.18  0.85724
## Year2012         -0.1043      0.1945     -0.54  0.59205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.144, Adjusted R-squared:  0.108
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 499 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.212  0.851  0.947  0.897  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          1.83e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.390 1          1.179
## LastAuthorFemale  1.230 1          1.109
## Year              1.314 16          1.009

```

Residuals from first and last author



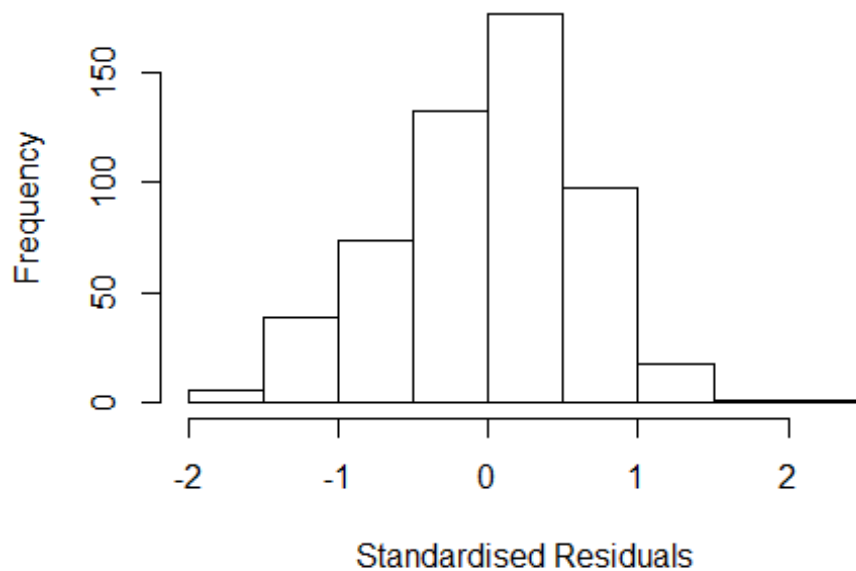
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7923 -0.4202 0.0525 0.4057 2.0775
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19596 0.20057 5.96 4.5e-09 ***
## FirstAuthorFemale1 0.13882 0.07175 1.93 0.054 .
## LastAuthorFemale1 0.02026 0.07487 0.27 0.787
## Year1997 -0.10469 0.30643 -0.34 0.733
## Year1998 0.35819 0.25667 1.40 0.163
## Year1999 0.27528 0.23645 1.16 0.245
## Year2000 0.27511 0.26856 1.02 0.306
## Year2001 0.45752 0.26442 1.73 0.084 .
## Year2002 -0.35834 0.25642 -1.40 0.163
## Year2003 -0.16141 0.24919 -0.65 0.517
## Year2004 0.12496 0.22849 0.55 0.585
## Year2005 -0.38231 0.23361 -1.64 0.102
```

```

## Year2006          0.03868    0.22201    0.17    0.862
## Year2007         -0.04826    0.23748   -0.20    0.839
## Year2008          0.32229    0.22071    1.46    0.145
## Year2009          0.15846    0.22034    0.72    0.472
## Year2010         -0.00613    0.21347   -0.03    0.977
## Year2011          0.07420    0.21551    0.34    0.731
## Year2012          0.02254    0.21501    0.10    0.917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0731
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.220  0.852  0.945  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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.223 1      1.106
## Year              1.223 16      1.006

```

Residuals from first author



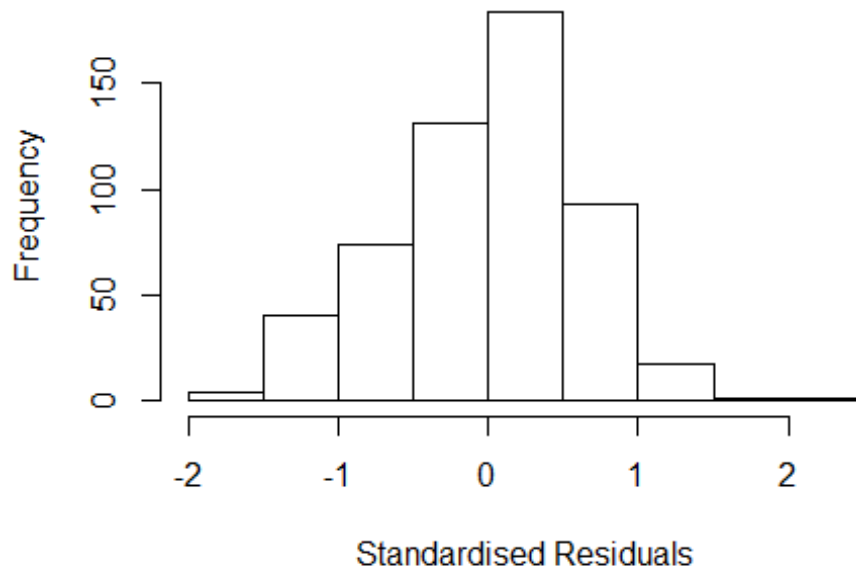
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8011 -0.4230 0.0589 0.4033 2.0760
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19460 0.19996 5.97 4.3e-09 ***
## FirstAuthorFemale1 0.14688 0.06744 2.18 0.030 *
## Year1997 -0.10504 0.30619 -0.34 0.732
## Year1998 0.36438 0.25506 1.43 0.154
## Year1999 0.28037 0.23448 1.20 0.232
## Year2000 0.27912 0.26719 1.04 0.297
## Year2001 0.45967 0.26387 1.74 0.082 .
## Year2002 -0.35556 0.25294 -1.41 0.160
## Year2003 -0.15990 0.24769 -0.65 0.519
## Year2004 0.12737 0.22739 0.56 0.576
## Year2005 -0.37720 0.23246 -1.62 0.105
## Year2006 0.04261 0.22092 0.19 0.847
```

```

## Year2007      -0.04545    0.23673   -0.19    0.848
## Year2008      0.32607    0.21950    1.49    0.138
## Year2009      0.16117    0.21947    0.73    0.463
## Year2010     -0.00262    0.21218   -0.01    0.990
## Year2011      0.07515    0.21498    0.35    0.727
## Year2012      0.02537    0.21423    0.12    0.906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0748
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 497 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.221  0.852  0.945  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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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
## -1.6474 -0.4422 0.0396 0.4296 2.0489
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21700 0.20675 5.89 7e-09 ***
## LastAuthorFemale1 0.08802 0.07077 1.24 0.21
## Year1997 -0.09642 0.31017 -0.31 0.76
## Year1998 0.32433 0.25797 1.26 0.21
## Year1999 0.24726 0.24049 1.03 0.30
## Year2000 0.26361 0.27473 0.96 0.34
## Year2001 0.43036 0.26570 1.62 0.11
## Year2002 -0.36893 0.25980 -1.42 0.16
## Year2003 -0.16393 0.25310 -0.65 0.52
## Year2004 0.15469 0.23411 0.66 0.51
## Year2005 -0.39030 0.23908 -1.63 0.10
## Year2006 0.03654 0.22859 0.16 0.87
```

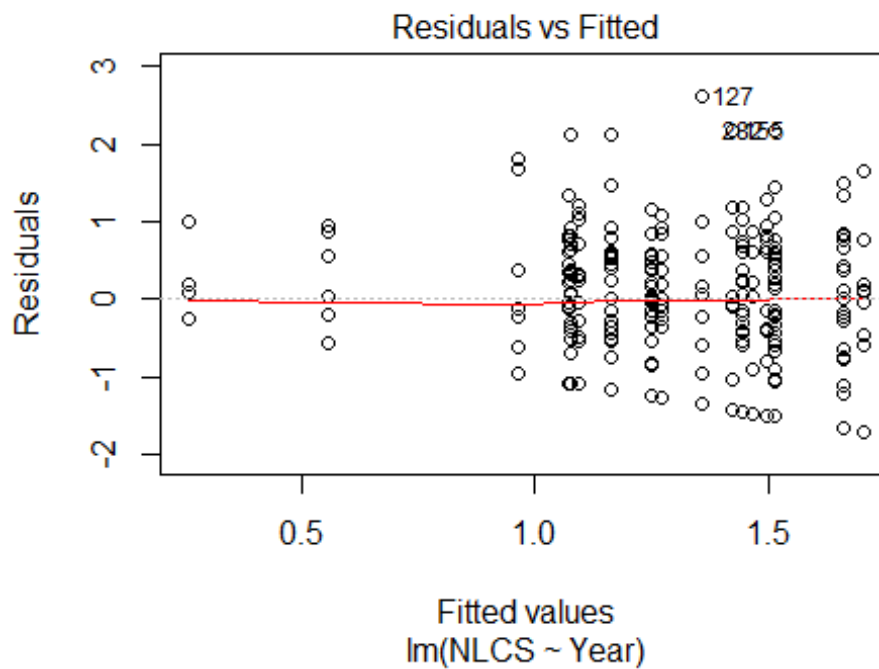


```

## Year2007          -0.04859      0.24373    -0.20      0.84
## Year2008           0.31872      0.22760      1.40      0.16
## Year2009           0.15836      0.22641      0.70      0.48
## Year2010          -0.00578      0.21976     -0.03      0.98
## Year2011           0.08942      0.22201      0.40      0.69
## Year2012           0.03013      0.22275      0.14      0.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.62
## Multiple R-squared:  0.0959, Adjusted R-squared:  0.0668
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 57 weights are ~= 1. The remaining 488 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.252  0.857   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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 545"
## [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
##   13   15   16   20   18   18   12   15   18   20   28   25   38   35   32
## 2011 2012
##   28   34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10    8    9   15    9   13    7   11   11   14   20   20   28   30   24
## 2011 2012

```

```
## 19 26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 7 9 13 6 10 7 11 11 13 19 18 27 26 24
## 2011 2012
## 15 21
## [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
```



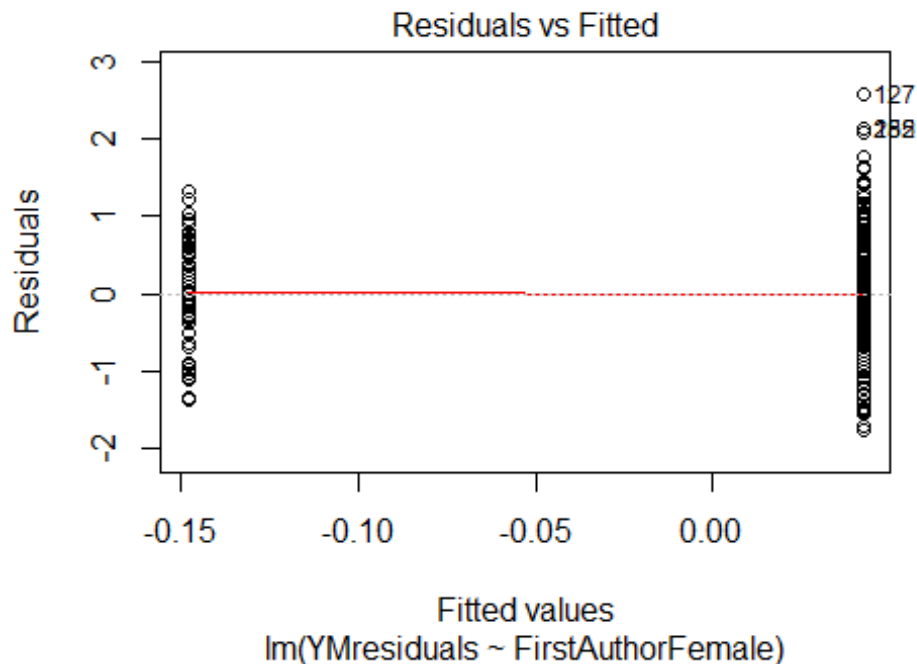
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.4, df = 1, p-value = 0.07

## [1] "Female first author team size 2018 geometric mean: 1.70673683684508"
## [1] "Male first author team size 2018 geometric mean: 2.04458356374137"

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

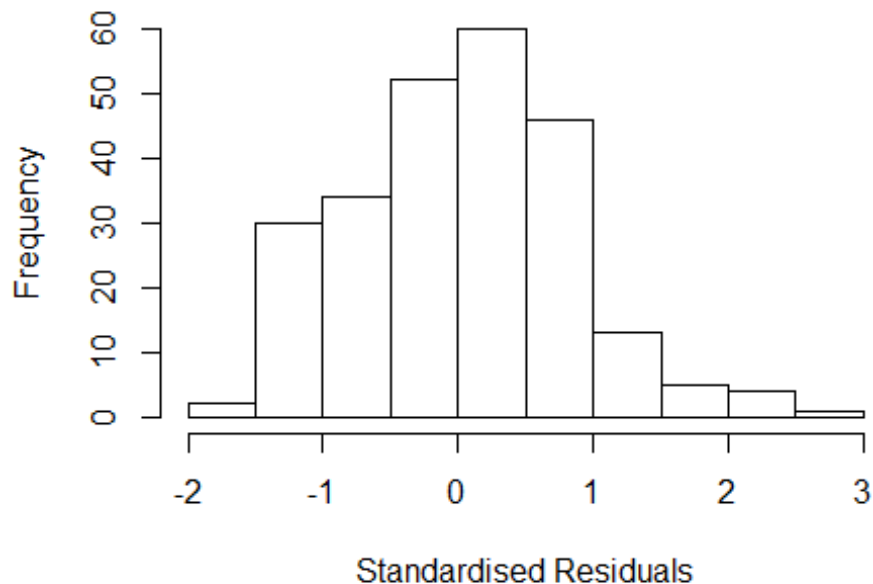
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.38071307156624"
## [1] "Male last author team size 2018 geometric mean: 2.16433134621535"

## 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 = 0.02
## 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 1.829  1      1.352
## LastAuthorFemale  1.790  1      1.338
## UniqueAuthors    3.846  3      1.252
## Year              6.109 16      1.058
```

Residuals from first and last author and team size



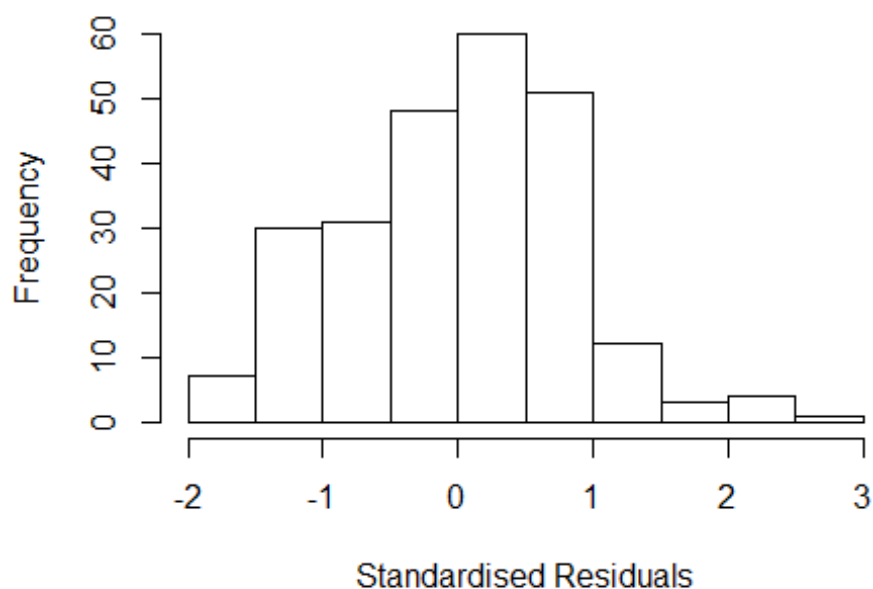
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 127 33747110111 3.977 2003      1213      2      2.627
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6348 -0.5431  0.0332  0.5688  2.6272
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5431    0.1901    2.86  0.0047 **
## FirstAuthorFemale1 -0.3464    0.1482   -2.34  0.0203 *
## LastAuthorFemale1  0.2789    0.1618    1.72  0.0861 .
## UniqueAuthors2     0.0896    0.1593    0.56  0.5742
## UniqueAuthors3     0.2548    0.1601    1.59  0.1129
## UniqueAuthors4     0.2349    0.1875    1.25  0.2116
## Year1997          -0.2670    0.2546   -1.05  0.2955
## Year1998           0.2842    0.4042    0.70  0.4828
## Year1999           0.3716    0.2868    1.30  0.1965
## Year2000           0.9560    0.4898    1.95  0.0522 .
```

```

## Year2001          0.9246      0.4354      2.12      0.0348 *
## Year2002          0.9397      0.4334      2.17      0.0312 *
## Year2003          0.5519      0.3458      1.60      0.1118
## Year2004          0.7041      0.5662      1.24      0.2149
## Year2005          0.4844      0.3144      1.54      0.1248
## Year2006          0.6273      0.2829      2.22      0.0276 *
## Year2007          0.8661      0.2888      3.00      0.0030 **
## Year2008          0.8794      0.2882      3.05      0.0026 **
## Year2009          0.6669      0.2334      2.86      0.0047 **
## Year2010          0.5030      0.2652      1.90      0.0592 .
## Year2011          1.0917      0.3407      3.20      0.0016 **
## Year2012          0.3728      0.2721      1.37      0.1721
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.844
## Multiple R-squared:  0.141, Adjusted R-squared:  0.061
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 220 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.312  0.879   0.951   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      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 1.732 1      1.316
## LastAuthorFemale  1.757 1      1.326
## Year              1.706 16      1.017

```

Residuals from first and last author



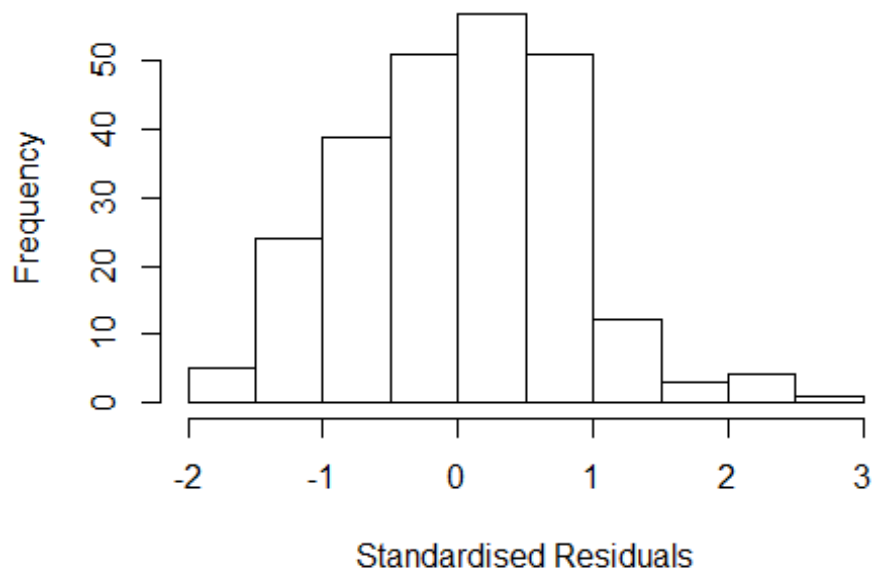
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 127 33747110111 3.977 2003      1213      2      2.808
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7434 -0.5602  0.0383  0.6001  2.8083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.543      0.190    2.85  0.00476 **
## FirstAuthorFemale1 -0.315      0.151   -2.09  0.03787 *
## LastAuthorFemale1  0.238      0.167    1.42  0.15596
## Year1997         -0.252      0.251   -1.01  0.31519
## Year1998          0.283      0.424    0.67  0.50612
## Year1999          0.420      0.293    1.43  0.15336
## Year2000          1.016      0.475    2.14  0.03338 *
## Year2001          1.046      0.408    2.56  0.01098 *
## Year2002          1.026      0.407    2.52  0.01237 *
## Year2003          0.626      0.340    1.84  0.06702 .
## Year2004          0.780      0.541    1.44  0.15069
## Year2005          0.540      0.307    1.76  0.07994 .
```

```

## Year2006          0.698      0.275      2.54  0.01175 *
## Year2007          0.961      0.268      3.58  0.00041 ***
## Year2008          0.996      0.258      3.86  0.00015 ***
## Year2009          0.782      0.226      3.45  0.00066 ***
## Year2010          0.552      0.256      2.15  0.03228 *
## Year2011          1.201      0.318      3.77  0.00021 ***
## Year2012          0.472      0.267      1.77  0.07781 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.834
## Multiple R-squared:  0.136, Adjusted R-squared:  0.0677
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 222 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.234  0.872  0.947  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      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 1.273 1      1.128
## Year              1.273 16      1.008

```

Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 127 33747110111 3.977 2003      1213      2      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
## -1.7115 -0.5647  0.0374  0.5908  2.7797
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.544      0.190    2.87  0.00455 **
## FirstAuthorFemale1 -0.180      0.131   -1.37  0.17110
## Year1997         -0.237      0.252   -0.94  0.34872
## Year1998          0.310      0.425    0.73  0.46677
## Year1999          0.448      0.300    1.49  0.13632
## Year2000          1.101      0.468    2.35  0.01955 *
## Year2001          1.035      0.412    2.51  0.01277 *
## Year2002          1.006      0.395    2.55  0.01147 *
## Year2003          0.654      0.356    1.84  0.06727 .
## Year2004          0.809      0.510    1.59  0.11373
## Year2005          0.559      0.297    1.88  0.06115 .
## Year2006          0.729      0.270    2.71  0.00732 **
```

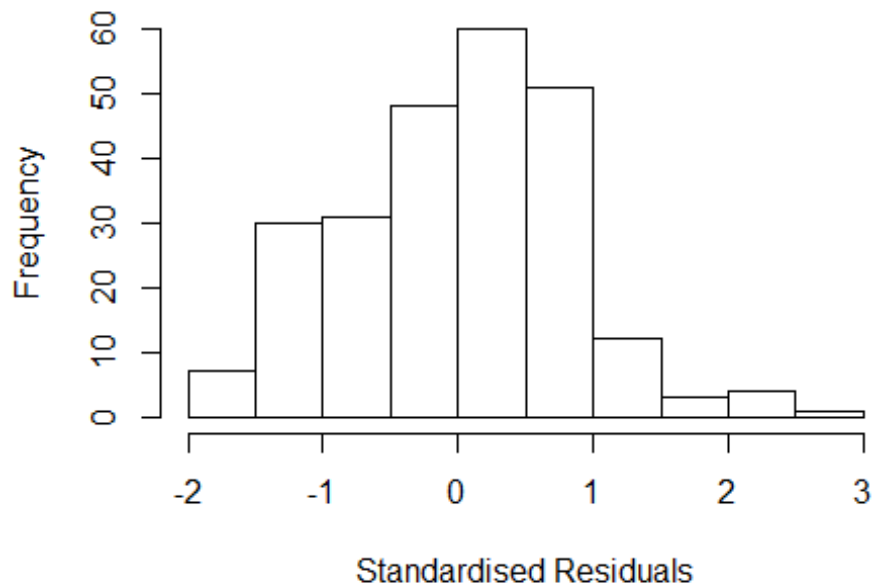


```

## Year2007          0.942      0.267      3.53  0.00050 ***
## Year2008          1.026      0.261      3.93  0.00011 ***
## Year2009          0.782      0.226      3.46  0.00064 ***
## Year2010          0.584      0.255      2.29  0.02269 *
## Year2011          1.168      0.318      3.67  0.00030 ***
## Year2012          0.472      0.267      1.77  0.07829 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.858
## Multiple R-squared:  0.128, Adjusted R-squared:  0.0633
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.273  0.888  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      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 1.295 1      1.138
## Year            1.295 16      1.008

```

Residuals from last author



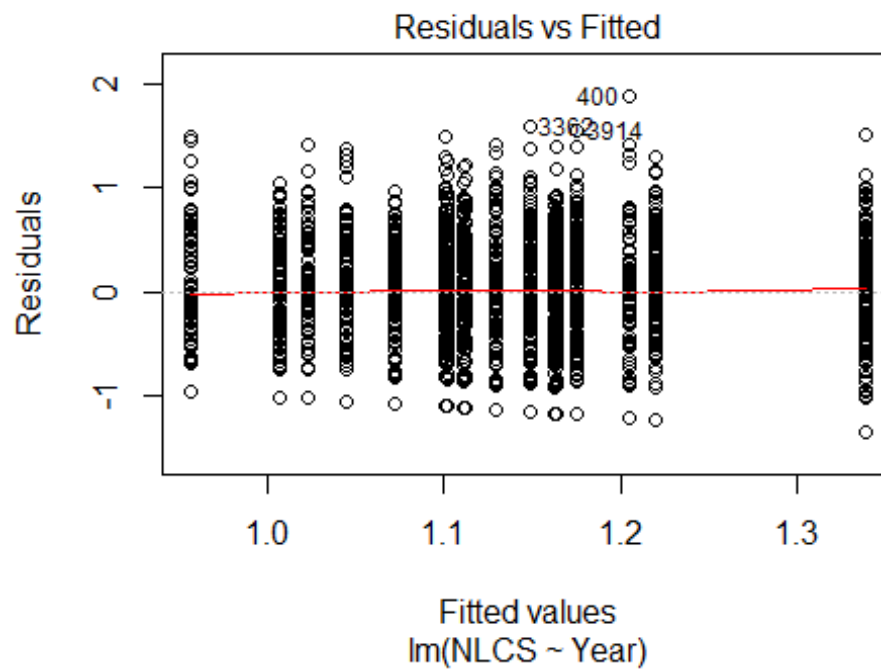
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 127 33747110111 3.977 2003      1213      2      2.808
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.601 -0.543  0.038  0.574  2.789
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5430     0.1902   2.86 0.00470 **
## LastAuthorFemale1 0.0299     0.1481   0.20 0.84037
## Year1997      -0.2687     0.2491  -1.08 0.28179
## Year1998       0.2670     0.4208   0.63 0.52632
## Year1999       0.4274     0.2961   1.44 0.15029
## Year2000       1.0415     0.4768   2.18 0.02994 *
## Year2001       1.0312     0.4276   2.41 0.01668 *
## Year2002       0.9857     0.3827   2.58 0.01063 *
## Year2003       0.6452     0.3522   1.83 0.06827 .
## Year2004       0.8023     0.5164   1.55 0.12165
## Year2005       0.5203     0.3165   1.64 0.10160
## Year2006       0.6905     0.2726   2.53 0.01196 *
```

```

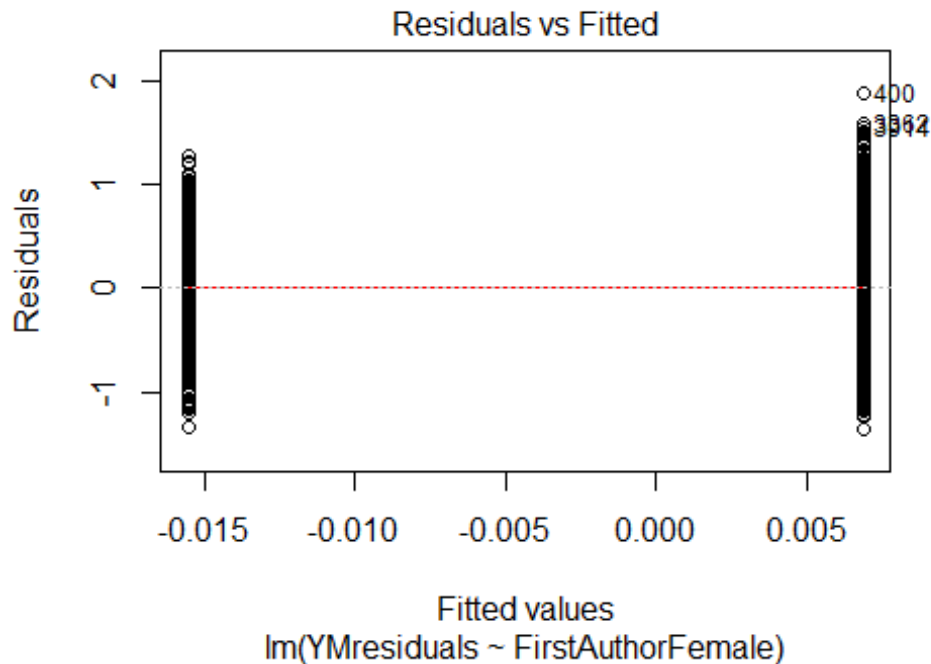
## Year2007          0.8865      0.2596      3.42  0.00075 ***
## Year2008          0.9498      0.2627      3.62  0.00037 ***
## Year2009          0.7322      0.2254      3.25  0.00133 **
## Year2010          0.5002      0.2539      1.97  0.05005 .
## Year2011          1.0578      0.3118      3.39  0.00081 ***
## Year2012          0.4188      0.2612      1.60  0.11017
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.84
## Multiple R-squared:  0.124, Adjusted R-squared:  0.0588
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 223 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.248  0.877  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      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 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
## 201 167 197 154 177 211 176 158 220 223 256 236 255 275 241
## 2011 2012
## 208 260
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 88 71 98 93 80 84 122 101 131 136 159 158 158 187 159
## 2011 2012

```

```
## 132 166
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 84 62 89 87 70 75 112 94 116 121 138 145 138 166 140
## 2011 2012
## 117 152
## [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
```

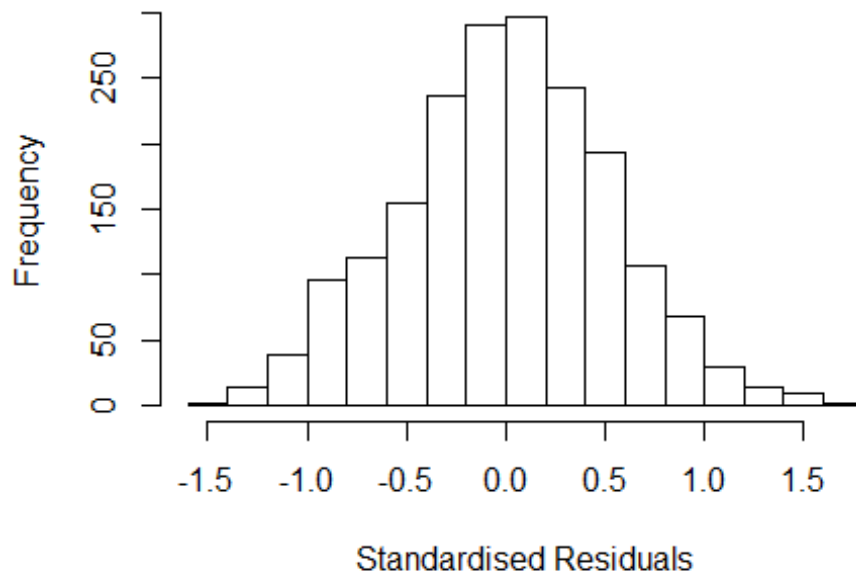


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.1, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 3.08921484123911"
## [1] "Male first author team size 2018 geometric mean: 2.64759249633892"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5100, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.65370389675083"
## [1] "Male last author team size 2018 geometric mean: 2.84504183012727"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, 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 1.225 1          1.107
## LastAuthorFemale  1.247 1          1.117
## UniqueAuthors     1.316 4          1.035
## Year              1.353 16          1.010
```

Residuals from first and last author and team size



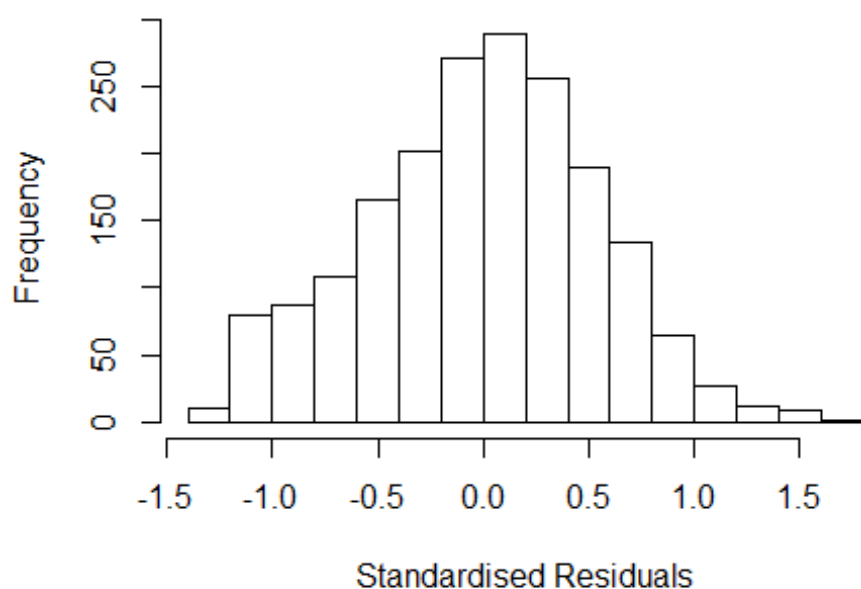
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47858 -0.34438  0.00481  0.35853  1.71794
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96652    0.07607   12.71 < 2e-16 ***
## FirstAuthorFemale1 0.01655    0.02861    0.58  0.563
## LastAuthorFemale1 -0.06221    0.03114   -2.00  0.046 *
## UniqueAuthors2    0.18587    0.03468    5.36 9.3e-08 ***
## UniqueAuthors3    0.24894    0.03755    6.63 4.4e-11 ***
## UniqueAuthors4    0.33612    0.04324    7.77 1.2e-14 ***
## UniqueAuthors5    0.41411    0.04297    9.64 < 2e-16 ***
## Year1997          0.14312    0.10609    1.35  0.177
## Year1998         -0.14375    0.09813   -1.46  0.143
## Year1999          0.02347    0.10104    0.23  0.816
```

```

## Year2000      -0.04365      0.09922      -0.44      0.660
## Year2001      -0.13494      0.09993      -1.35      0.177
## Year2002      -0.01368      0.09408      -0.15      0.884
## Year2003       0.02998      0.08804       0.34      0.733
## Year2004      -0.12934      0.08914      -1.45      0.147
## Year2005      -0.04729      0.08385      -0.56      0.573
## Year2006      -0.05531      0.08586      -0.64      0.520
## Year2007      -0.01508      0.08495      -0.18      0.859
## Year2008       0.00933      0.08730       0.11      0.915
## Year2009      -0.00212      0.08874      -0.02      0.981
## Year2010       0.09796      0.08834       1.11      0.268
## Year2011      -0.00819      0.09314      -0.09      0.930
## Year2012       0.17355      0.08950       1.94      0.053 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.0904, Adjusted R-squared:  0.0798
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 1722 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  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      5.25e-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 1.197 1      1.094
## LastAuthorFemale  1.201 1      1.096
## 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
## -1.3899 -0.3623 0.0191 0.3690 1.6853
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0717 0.0807 13.28 <2e-16 ***
## FirstAuthorFemale1 0.0264 0.0294 0.90 0.3694
## LastAuthorFemale1 -0.0870 0.0318 -2.73 0.0063 **
## Year1997 0.1179 0.1138 1.04 0.3003
## Year1998 -0.1376 0.1029 -1.34 0.1814
## Year1999 0.0551 0.1087 0.51 0.6127
## Year2000 0.0126 0.1037 0.12 0.9032
## Year2001 -0.0958 0.1066 -0.90 0.3689
## Year2002 0.0692 0.0999 0.69 0.4889
## Year2003 0.1012 0.0927 1.09 0.2750
## Year2004 -0.0721 0.0938 -0.77 0.4420
## Year2005 0.0294 0.0886 0.33 0.7402
```

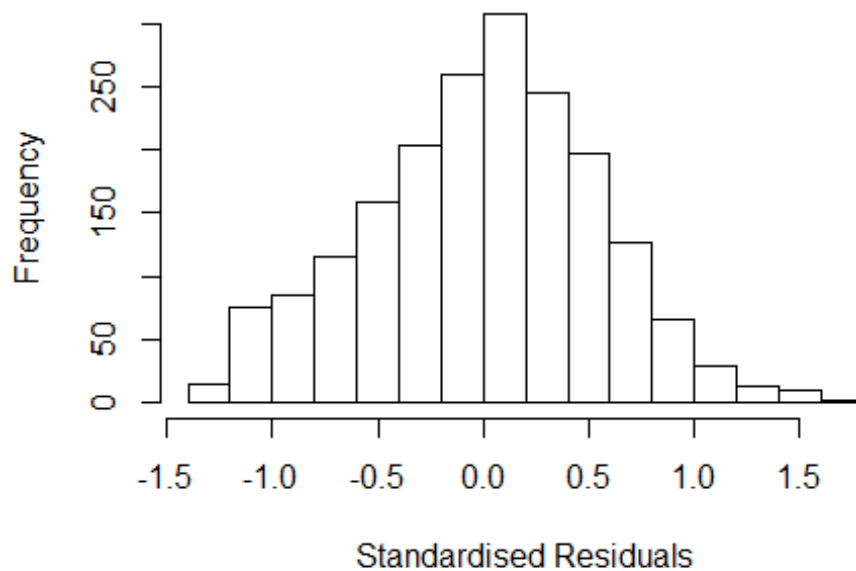


```

## Year2006          0.0283      0.0902      0.31      0.7539
## Year2007          0.0699      0.0894      0.78      0.4342
## Year2008          0.1058      0.0917      1.15      0.2485
## Year2009          0.0780      0.0943      0.83      0.4085
## Year2010          0.1867      0.0928      2.01      0.0443 *
## Year2011          0.1057      0.0996      1.06      0.2887
## Year2012          0.2919      0.0939      3.11      0.0019 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0357, Adjusted R-squared:  0.0265
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1732 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.316  0.868  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      5.25e-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 1.027 1      1.013
## Year              1.027 16      1.001

```

Residuals from first author



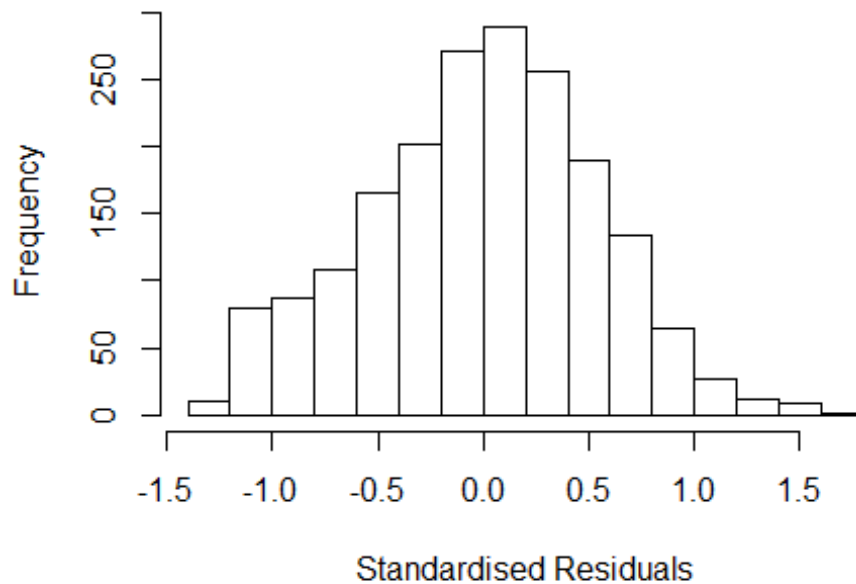
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3532 -0.3646 0.0228 0.3691 1.6070
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05440 0.07907 13.33 <2e-16 ***
## FirstAuthorFemale1 -0.00561 0.02744 -0.20 0.8381
## Year1997 0.12594 0.11397 1.11 0.2693
## Year1998 -0.13236 0.10259 -1.29 0.1971
## Year1999 0.05898 0.10857 0.54 0.5870
## Year2000 0.02147 0.10268 0.21 0.8344
## Year2001 -0.08906 0.10489 -0.85 0.3960
## Year2002 0.07871 0.09883 0.80 0.4259
## Year2003 0.11087 0.09146 1.21 0.2256
## Year2004 -0.06205 0.09269 -0.67 0.5033
## Year2005 0.03411 0.08736 0.39 0.6962
## Year2006 0.03755 0.08904 0.42 0.6733
```

```

## Year2007          0.07234    0.08843    0.82    0.4134
## Year2008          0.10651    0.09057    1.18    0.2397
## Year2009          0.08661    0.09345    0.93    0.3542
## Year2010          0.18527    0.09177    2.02    0.0437 *
## Year2011          0.10826    0.09947    1.09    0.2766
## Year2012          0.29880    0.09274    3.22    0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0317, Adjusted R-squared:  0.023
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.362  0.869  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      5.25e-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      1.016
## Year              1.031 16      1.001

```

Residuals from last author



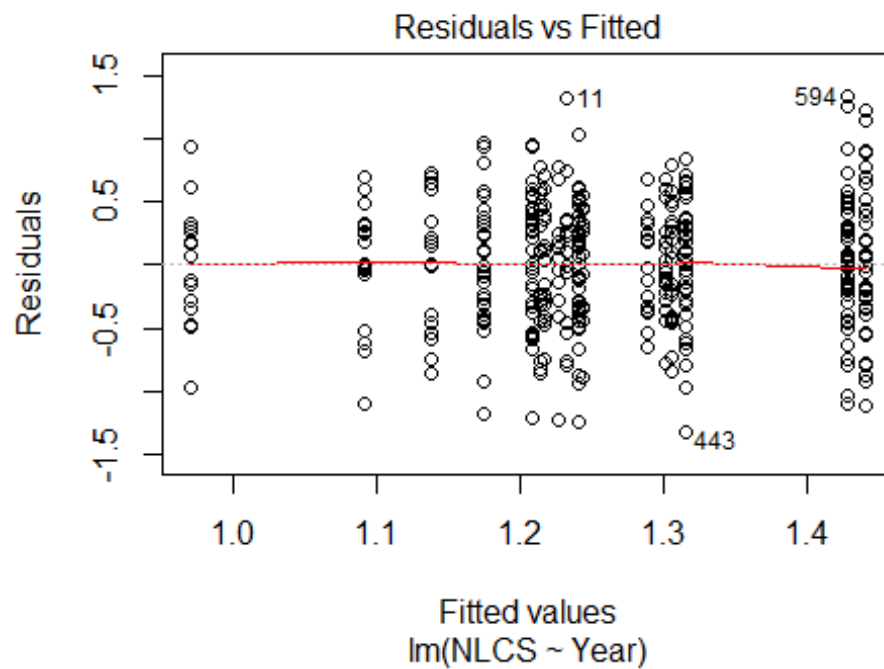
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3698 -0.3633 0.0199 0.3740 1.6688
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0752 0.0807 13.33 <2e-16 ***
## LastAuthorFemale1 -0.0757 0.0294 -2.57 0.0102 *
## Year1997 0.1188 0.1140 1.04 0.2978
## Year1998 -0.1358 0.1028 -1.32 0.1865
## Year1999 0.0560 0.1086 0.52 0.6065
## Year2000 0.0139 0.1036 0.13 0.8935
## Year2001 -0.0922 0.1063 -0.87 0.3856
## Year2002 0.0698 0.0999 0.70 0.4850
## Year2003 0.1034 0.0926 1.12 0.2645
## Year2004 -0.0690 0.0937 -0.74 0.4620
## Year2005 0.0325 0.0884 0.37 0.7133
## Year2006 0.0316 0.0901 0.35 0.7256
```

```

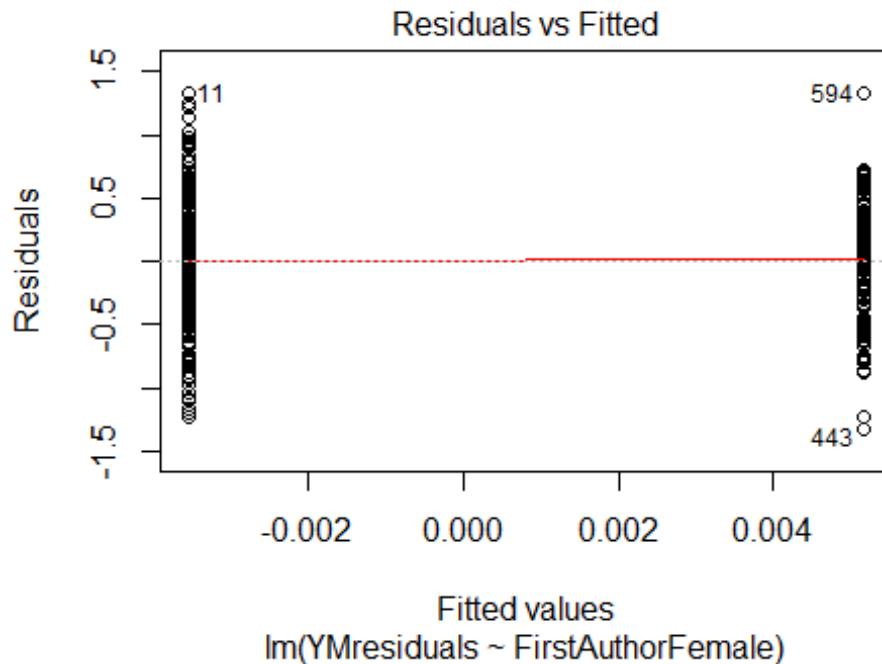
## Year2007          0.0708      0.0894      0.79      0.4283
## Year2008          0.1083      0.0915      1.18      0.2370
## Year2009          0.0798      0.0943      0.85      0.3979
## Year2010          0.1889      0.0927      2.04      0.0418 *
## Year2011          0.1083      0.0998      1.09      0.2780
## Year2012          0.2946      0.0939      3.14      0.0017 **
## ---
## 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:  0.0266
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.326  0.868  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      5.25e-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: 1906"
## [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
##   23   22   26   30   31   30   25   18   19   23   33   21   38   46   41
## 2011 2012
##   77   69
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   16   13   18   16   11   20   15   16   19   31   20   35   38   35
## 2011 2012

```

```
## 44 48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 16 13 18 16 10 19 14 16 19 26 18 33 36 32
## 2011 2012
## 40 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 = 12, df = 16, p-value = 0.8
```

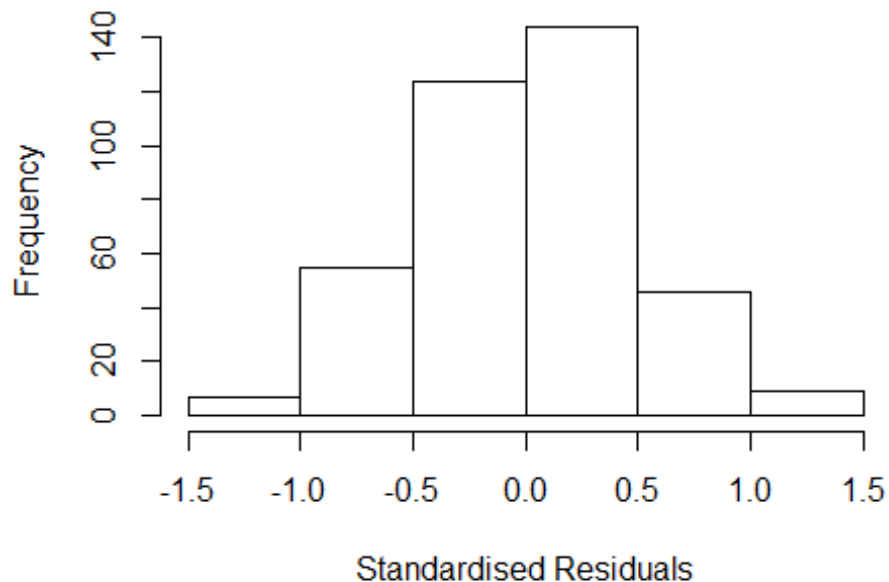


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.3, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 2.29343777445726"
## [1] "Male first author team size 2018 geometric mean: 2.53570235181859"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.09017533402798"
## [1] "Male last author team size 2018 geometric mean: 2.70721068805566"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 870, 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 2.092 1      1.446
## LastAuthorFemale  2.435 1      1.560
## UniqueAuthors    3.456 4      1.168
## Year              3.694 16     1.042
```

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
## -1.4358 -0.3408 0.0236 0.3142 1.4333
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.182761 0.181149 6.53 2.2e-10 ***
## FirstAuthorFemale1 0.013657 0.070492 0.19 0.846
## LastAuthorFemale1 0.008985 0.074728 0.12 0.904
## UniqueAuthors2 0.160787 0.070937 2.27 0.024 *
## UniqueAuthors3 0.126671 0.085741 1.48 0.140
## UniqueAuthors4 0.250114 0.163827 1.53 0.128
## UniqueAuthors5 -0.000454 0.169538 0.00 0.998
## Year1997 0.021336 0.235443 0.09 0.928
## Year1998 0.010915 0.222332 0.05 0.961
## Year1999 -0.024793 0.206773 -0.12 0.905
```

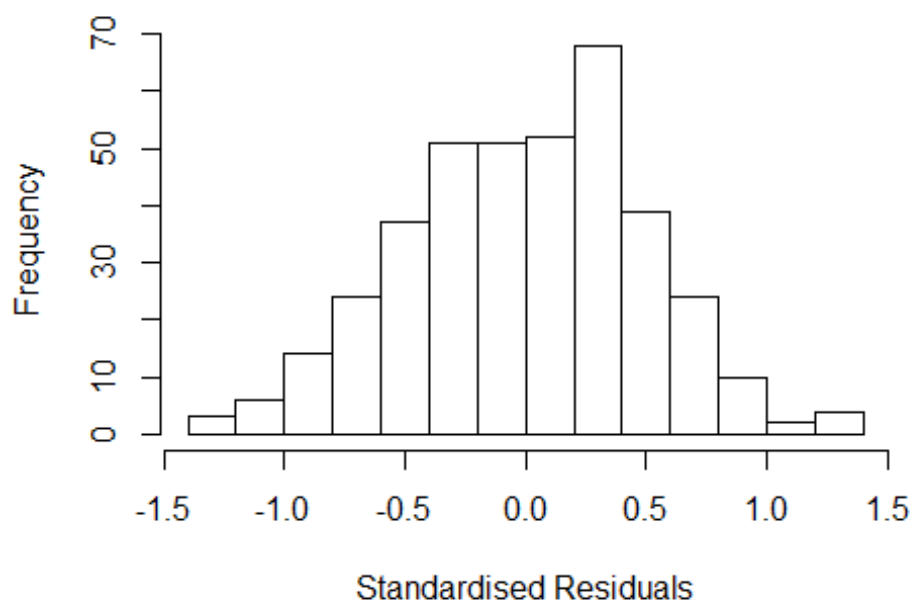


```

## Year2000      0.054685    0.205889    0.27    0.791
## Year2001     -0.075724    0.222299   -0.34    0.734
## Year2002     -0.107734    0.214095   -0.50    0.615
## Year2003     -0.341142    0.219377   -1.56    0.121
## Year2004     -0.095978    0.235798   -0.41    0.684
## Year2005      0.027957    0.202454    0.14    0.890
## Year2006     -0.053309    0.211980   -0.25    0.802
## Year2007      0.072195    0.212383    0.34    0.734
## Year2008     -0.012199    0.208744   -0.06    0.953
## Year2009      0.078626    0.198551    0.40    0.692
## Year2010      0.164319    0.217438    0.76    0.450
## Year2011     -0.052653    0.205847   -0.26    0.798
## Year2012      0.130741    0.208574    0.63    0.531
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.493
## Multiple R-squared:  0.0684, Adjusted R-squared:  0.0117
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 347 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.377  0.873   0.953   0.906   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 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.987 1      1.410
## LastAuthorFemale  1.905 1      1.380
## Year              1.290 16      1.008

```

Residuals from first and last author



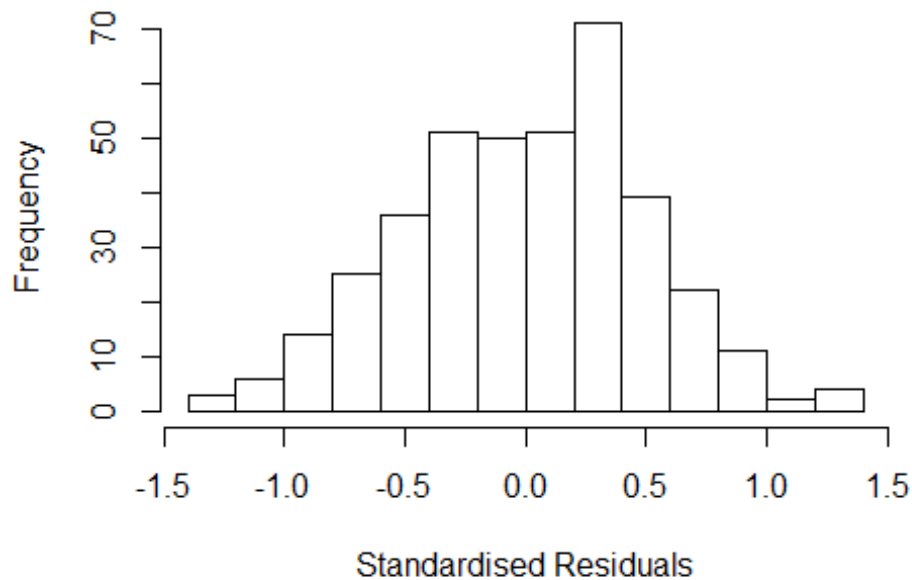
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3657 -0.3681 0.0224 0.3155 1.3554
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1966 0.1738 6.89 2.5e-11 ***
## FirstAuthorFemale1 0.0410 0.0702 0.58 0.56
## LastAuthorFemale1 -0.0353 0.0693 -0.51 0.61
## Year1997 0.0358 0.2269 0.16 0.87
## Year1998 0.0543 0.2123 0.26 0.80
## Year1999 0.0156 0.1970 0.08 0.94
## Year2000 0.0855 0.1988 0.43 0.67
## Year2001 -0.0175 0.2187 -0.08 0.94
## Year2002 -0.0827 0.2053 -0.40 0.69
## Year2003 -0.2771 0.2113 -1.31 0.19
## Year2004 -0.0579 0.2281 -0.25 0.80
## Year2005 0.0988 0.1915 0.52 0.61
```

```

## Year2006          -0.0278      0.2019    -0.14      0.89
## Year2007           0.1048      0.2095      0.50      0.62
## Year2008           0.0350      0.1982      0.18      0.86
## Year2009           0.1281      0.1910      0.67      0.50
## Year2010           0.2109      0.2085      1.01      0.31
## Year2011           0.0488      0.1896      0.26      0.80
## Year2012           0.2295      0.1902      1.21      0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0477, Adjusted R-squared:  0.000824
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 358 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.476  0.889  0.955  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      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 1.196 1      1.093
## Year              1.196 16      1.006

```

Residuals from first author

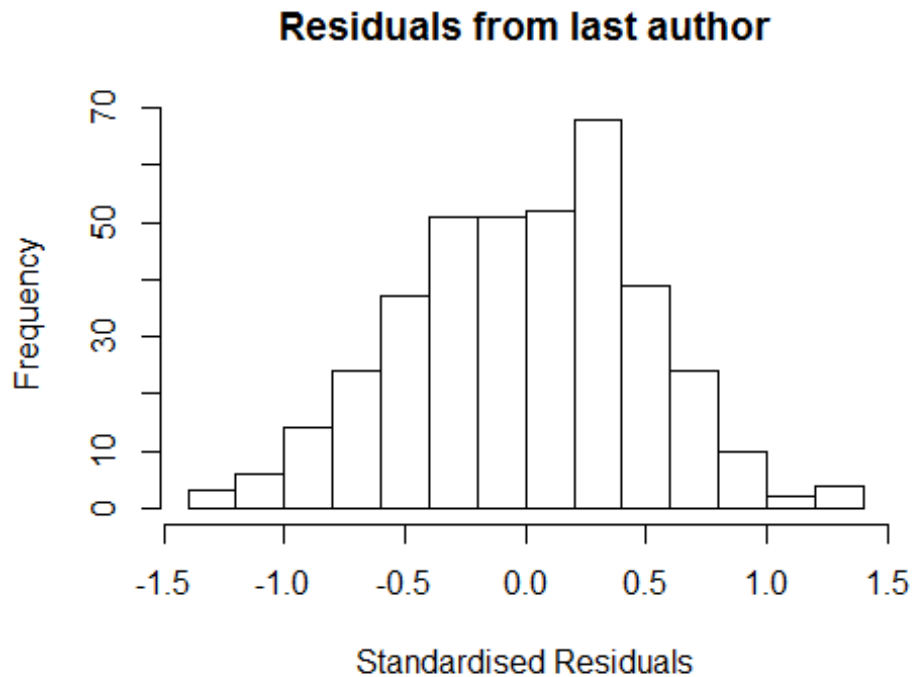


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3411 -0.3691  0.0251  0.3239  1.3581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1939     0.1723   6.93 1.9e-11 ***
## FirstAuthorFemale1  0.0190     0.0546   0.35  0.73
## Year1997          0.0332     0.2271   0.15  0.88
## Year1998          0.0522     0.2115   0.25  0.81
## Year1999          0.0153     0.1967   0.08  0.94
## Year2000          0.0860     0.1973   0.44  0.66
## Year2001         -0.0158     0.2176  -0.07  0.94
## Year2002         -0.0868     0.2046  -0.42  0.67
## Year2003         -0.2771     0.2094  -1.32  0.19
## Year2004         -0.0551     0.2267  -0.24  0.81
## Year2005          0.0993     0.1899   0.52  0.60
## Year2006         -0.0240     0.2005  -0.12  0.90
```

```

## Year2007          0.1018      0.2094      0.49      0.63
## Year2008          0.0322      0.1982      0.16      0.87
## Year2009          0.1282      0.1902      0.67      0.50
## Year2010          0.2107      0.2080      1.01      0.31
## Year2011          0.0478      0.1889      0.25      0.80
## Year2012          0.2267      0.1888      1.20      0.23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0469, Adjusted R-squared:  0.00279
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 361 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.478  0.888  0.955  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      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 1.153 1      1.074
## Year            1.153 16      1.004

```



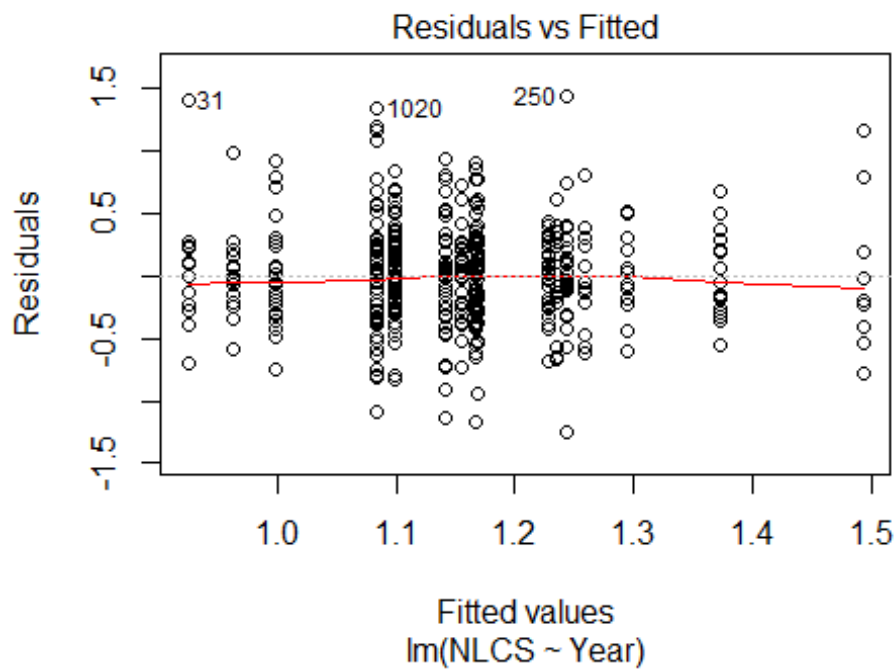
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3326 -0.3666  0.0296  0.3283  1.3494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20258    0.17233   6.98 1.4e-11 ***
## LastAuthorFemale1 -0.00838    0.05395  -0.16  0.88
## Year1997         0.02991    0.22587   0.13  0.89
## Year1998         0.05900    0.21169   0.28  0.78
## Year1999         0.01490    0.19678   0.08  0.94
## Year2000         0.09042    0.19753   0.46  0.65
## Year2001        -0.01457    0.21845  -0.07  0.95
## Year2002        -0.08059    0.20370  -0.40  0.69
## Year2003        -0.27528    0.20829  -1.32  0.19
## Year2004        -0.05688    0.22664  -0.25  0.80
## Year2005         0.09887    0.19028   0.52  0.60
## Year2006        -0.02280    0.20036  -0.11  0.91
```

```

## Year2007      0.10951    0.20884    0.52    0.60
## Year2008      0.03810    0.19683    0.19    0.85
## Year2009      0.13001    0.19019    0.68    0.49
## Year2010      0.21403    0.20701    1.03    0.30
## Year2011      0.05165    0.18807    0.27    0.78
## Year2012      0.22922    0.18851    1.22    0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0467, Adjusted R-squared:  0.00251
## 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.
##  0.488  0.888  0.955  0.917  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      2.60e-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 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
##   34   29   31   29   23   56   29   34   28   39   52   51   46   61   80
## 2011 2012
##  113  103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   13   16   11    9   25   15   18   12   26   28   28   24   29   52
## 2011 2012

```

```
## 70 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 12 15 11 9 23 15 18 12 22 24 24 20 26 47
## 2011 2012
## 57 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 = 34, df = 16, p-value = 0.006
```

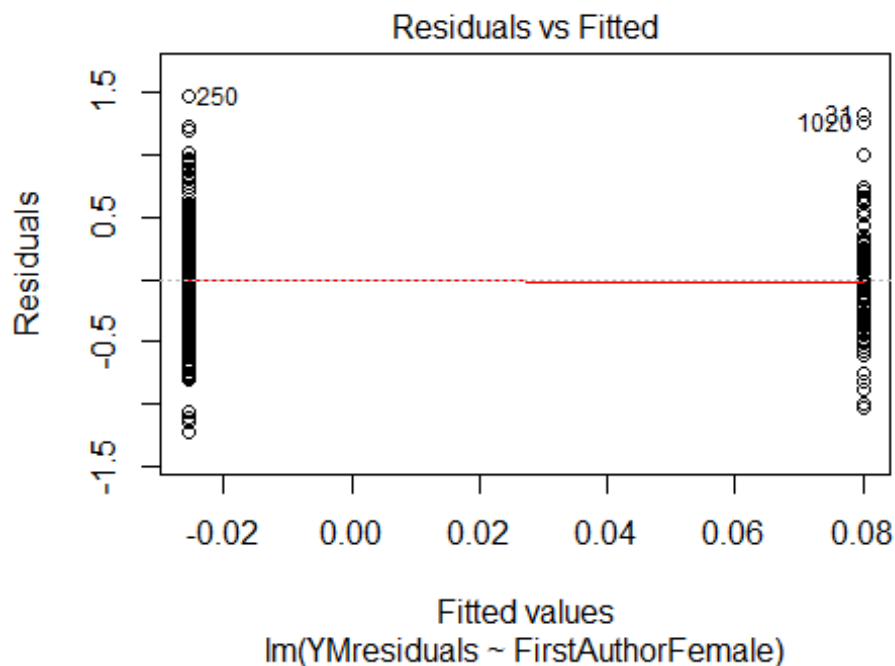


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.12, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 3.25557235911573"
## [1] "Male first author team size 2018 geometric mean: 4.31231009787138"
## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.45652939881143"
## [1] "Male last author team size 2018 geometric mean: 4.22752845081128"

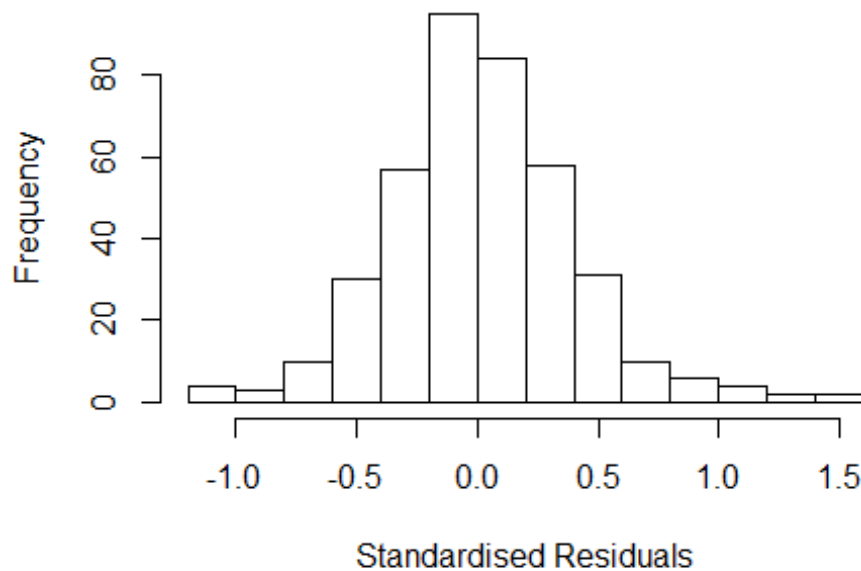
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 230, 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	1.319	1	1.149
LastAuthorFemale	1.138	1	1.067
UniqueAuthors	1.802	4	1.076
Year	2.293	16	1.026

Residuals from first and last author and team size



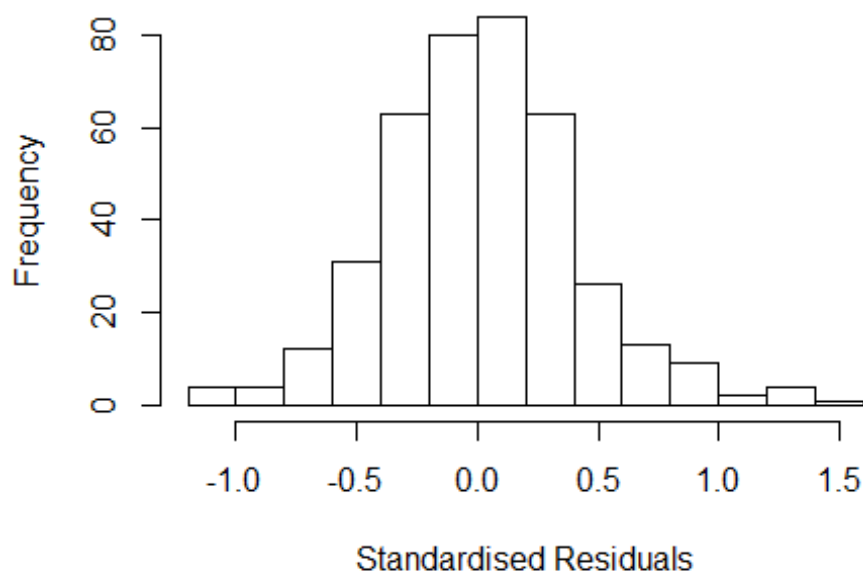
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11797 -0.21876 -0.00264 0.24315 1.46869
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.71896 0.09435 7.62 2.1e-13 ***
## FirstAuthorFemale1 0.09630 0.04795 2.01 0.04535 *
## LastAuthorFemale1 0.11932 0.05613 2.13 0.03419 *
## UniqueAuthors2 0.09923 0.06748 1.47 0.14230
## UniqueAuthors3 0.20187 0.06725 3.00 0.00287 **
## UniqueAuthors4 0.30571 0.07423 4.12 4.7e-05 ***
## UniqueAuthors5 0.32894 0.07291 4.51 8.6e-06 ***
## Year1997 0.41724 0.13375 3.12 0.00195 **
## Year1998 0.45500 0.13156 3.46 0.00061 ***
## Year1999 0.05693 0.27700 0.21 0.83726
```

```

## Year2000          0.40386    0.21106    1.91  0.05645 .
## Year2001          0.29978    0.11206    2.68  0.00780 **
## Year2002          0.31021    0.12507    2.48  0.01357 *
## Year2003         -0.00807    0.10855   -0.07  0.94076
## Year2004          0.35035    0.12768    2.74  0.00636 **
## Year2005          0.30024    0.11387    2.64  0.00872 **
## Year2006          0.23136    0.11221    2.06  0.03991 *
## Year2007          0.21293    0.11453    1.86  0.06381 .
## Year2008         -0.03677    0.11817   -0.31  0.75588
## Year2009          0.24456    0.10055    2.43  0.01548 *
## Year2010          0.21198    0.10251    2.07  0.03934 *
## Year2011          0.12021    0.09786    1.23  0.22007
## Year2012          0.03983    0.11135    0.36  0.72079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared:  0.2,   Adjusted R-squared:  0.153
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 367 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0447 0.8750 0.9540 0.8940 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          2.53e-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 1.309 1          1.144
## LastAuthorFemale 1.123 1          1.060
## Year              1.437 16          1.011

```

Residuals from first and last author



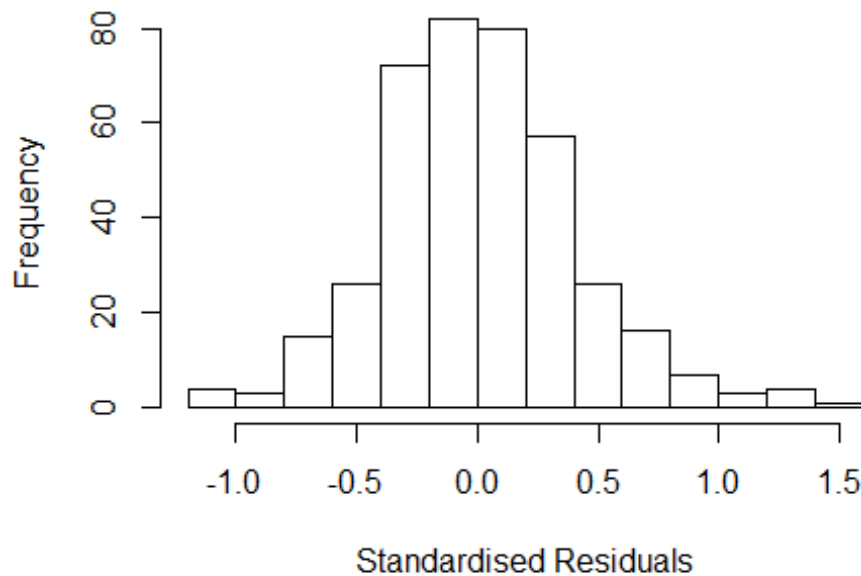
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17135 -0.22898 0.00559 0.23589 1.51165
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8402 0.0809 10.39 < 2e-16 ***
## FirstAuthorFemale1 0.1086 0.0501 2.17 0.03088 *
## LastAuthorFemale1 0.1238 0.0613 2.02 0.04401 *
## Year1997 0.3996 0.1462 2.73 0.00657 **
## Year1998 0.4728 0.1258 3.76 0.00020 ***
## Year1999 0.0906 0.2945 0.31 0.75837
## Year2000 0.4288 0.2031 2.11 0.03542 *
## Year2001 0.3311 0.1037 3.19 0.00153 **
## Year2002 0.3725 0.1352 2.75 0.00616 **
## Year2003 0.0493 0.1024 0.48 0.63081
## Year2004 0.4126 0.1290 3.20 0.00149 **
## Year2005 0.3111 0.1086 2.87 0.00440 **
```

```

## Year2006          0.2787      0.1092      2.55  0.01113 *
## Year2007          0.2843      0.1098      2.59  0.01000 **
## Year2008          0.0394      0.1166      0.34  0.73562
## Year2009          0.3429      0.0998      3.43  0.00066 ***
## Year2010          0.2914      0.1018      2.86  0.00446 **
## Year2011          0.2062      0.0955      2.16  0.03153 *
## Year2012          0.0968      0.1141      0.85  0.39686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.128, Adjusted R-squared:  0.0865
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 354 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0425 0.8640 0.9540 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      2.53e-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 1.272 1      1.128
## Year              1.272 16      1.008

```

Residuals from first author



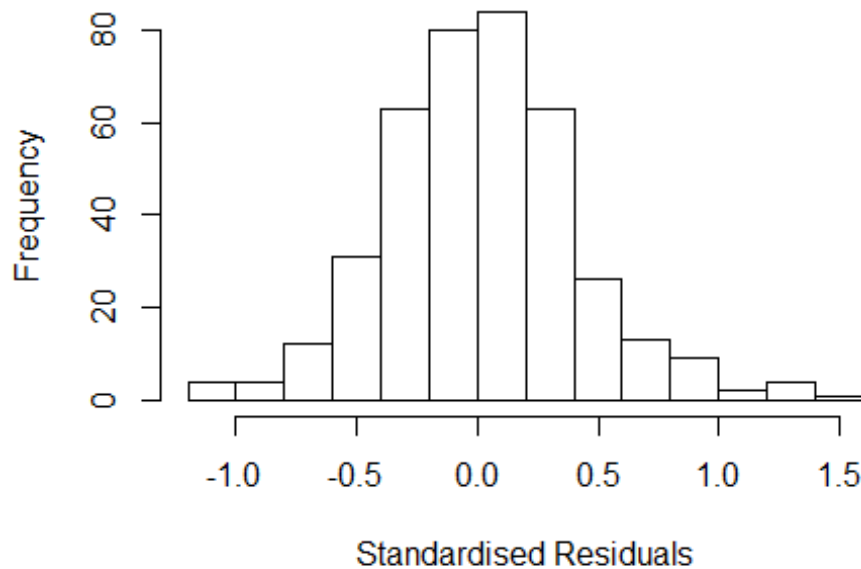
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19055 -0.23690 -0.00525  0.26066  1.49245
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8520    0.0842   10.12 < 2e-16 ***
## FirstAuthorFemale1 0.1202    0.0504    2.38  0.01763 *
## Year1997        0.3975    0.1513    2.63  0.00894 **
## Year1998        0.4596    0.1282    3.58  0.00038 ***
## Year1999        0.1446    0.2966    0.49  0.62629
## Year2000        0.4478    0.2053    2.18  0.02974 *
## Year2001        0.3385    0.1074    3.15  0.00175 **
## Year2002        0.3809    0.1433    2.66  0.00821 **
## Year2003        0.0337    0.1054    0.32  0.74976
## Year2004        0.4083    0.1347    3.03  0.00259 **
## Year2005        0.3066    0.1107    2.77  0.00590 **
## Year2006        0.2840    0.1122    2.53  0.01175 *
```

```

## Year2007          0.2911      0.1136      2.56  0.01075 *
## Year2008          0.0398      0.1197      0.33  0.74003
## Year2009          0.3541      0.1009      3.51  0.00050 ***
## Year2010          0.2914      0.1050      2.78  0.00577 **
## Year2011          0.2117      0.0984      2.15  0.03202 *
## Year2012          0.1199      0.1190      1.01  0.31426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.113, Adjusted R-squared:  0.0736
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 352 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.068  0.883   0.949   0.890   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      2.53e-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 1.107 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
## -1.18601 -0.23283 -0.00126 0.23255 1.49699
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8326 0.0785 10.61 < 2e-16 ***
## LastAuthorFemale1 0.1374 0.0619 2.22 0.02705 *
## Year1997 0.4275 0.1494 2.86 0.00446 **
## Year1998 0.5032 0.1208 4.17 3.9e-05 ***
## Year1999 0.1752 0.2748 0.64 0.52417
## Year2000 0.4628 0.2157 2.15 0.03250 *
## Year2001 0.3534 0.1022 3.46 0.00060 ***
## Year2002 0.3951 0.1309 3.02 0.00272 **
## Year2003 0.1038 0.0974 1.07 0.28730
## Year2004 0.4519 0.1253 3.61 0.00035 ***
## Year2005 0.3482 0.1056 3.30 0.00107 **
## Year2006 0.3098 0.1047 2.96 0.00328 **
```

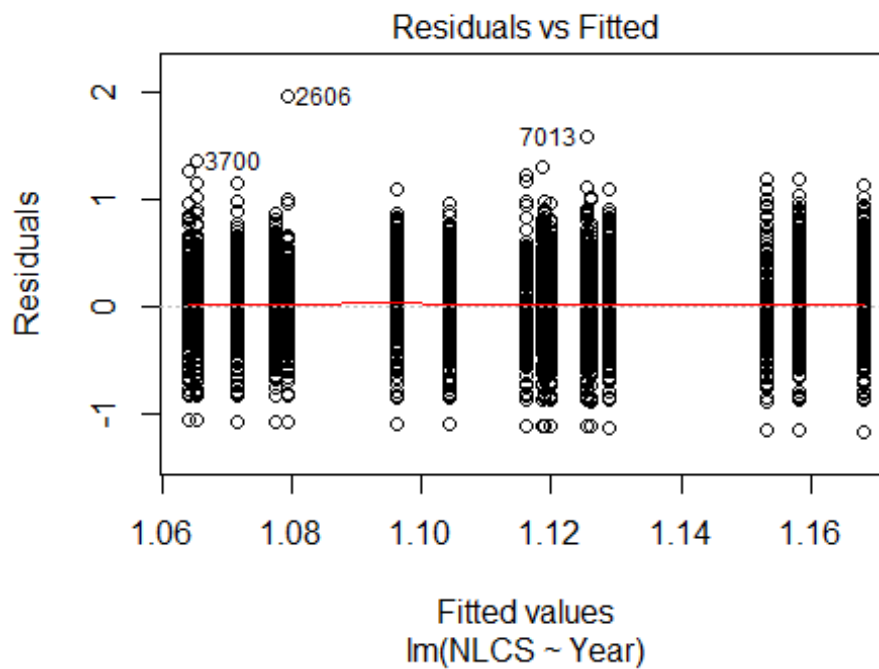


```

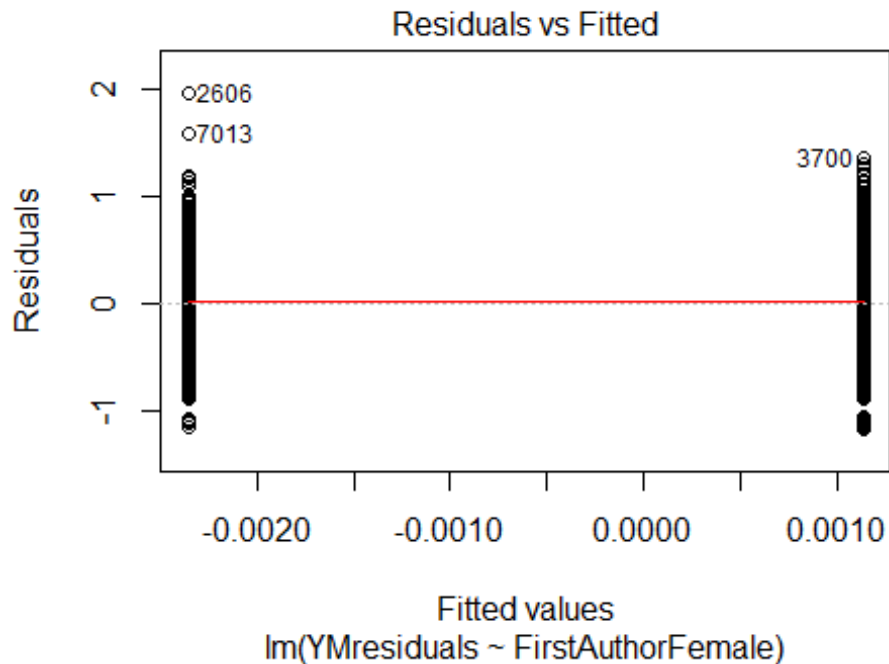
## Year2007          0.3041      0.1104      2.76  0.00614 **
## Year2008          0.0658      0.1119      0.59  0.55704
## Year2009          0.3806      0.0950      4.01  7.5e-05 ***
## Year2010          0.3184      0.0989      3.22  0.00140 **
## Year2011          0.2398      0.0922      2.60  0.00964 **
## Year2012          0.1220      0.1127      1.08  0.27992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.118, Adjusted R-squared:  0.0779
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 356 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0485 0.8700 0.9530 0.8860 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      2.53e-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: 396"
## [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
## 353 413 407 385 446 466 443 416 421 521 513 567 623 578 655
## 2011 2012
## 621 629
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 154 181 191 194 174 161 271 259 248 350 359 379 442 400 480
## 2011 2012

```

```
## 433 457
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 139 165 179 182 159 147 244 231 230 312 321 338 400 363 435
## 2011 2012
## 402 424
## [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
```

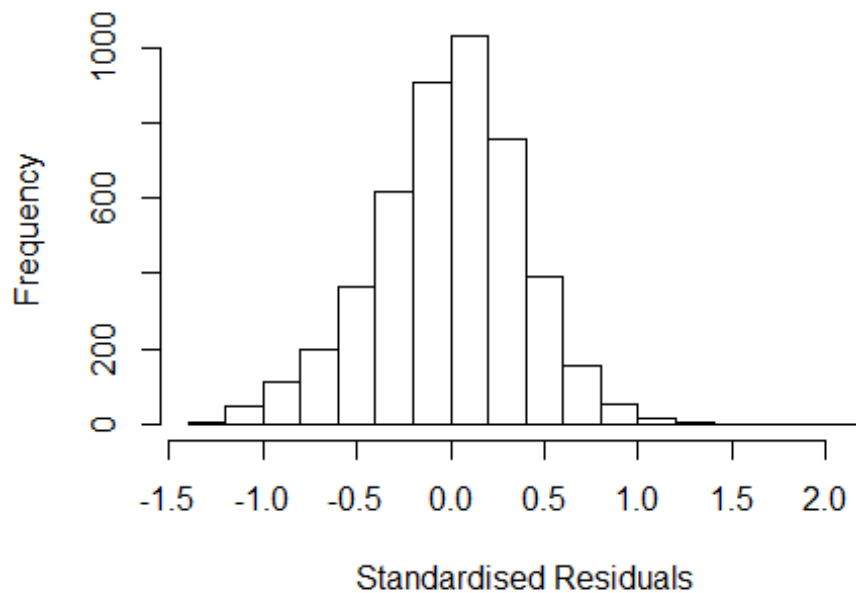


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.8, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.54751786088737"
## [1] "Male first author team size 2018 geometric mean: 3.49413095886021"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 29000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.42598102262956"
## [1] "Male last author team size 2018 geometric mean: 3.55206623186656"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, 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 1.087 1          1.043
## LastAuthorFemale  1.062 1          1.031
## UniqueAuthors    1.190 4          1.022
## Year              1.240 16         1.007
```

Residuals from first and last author and team size



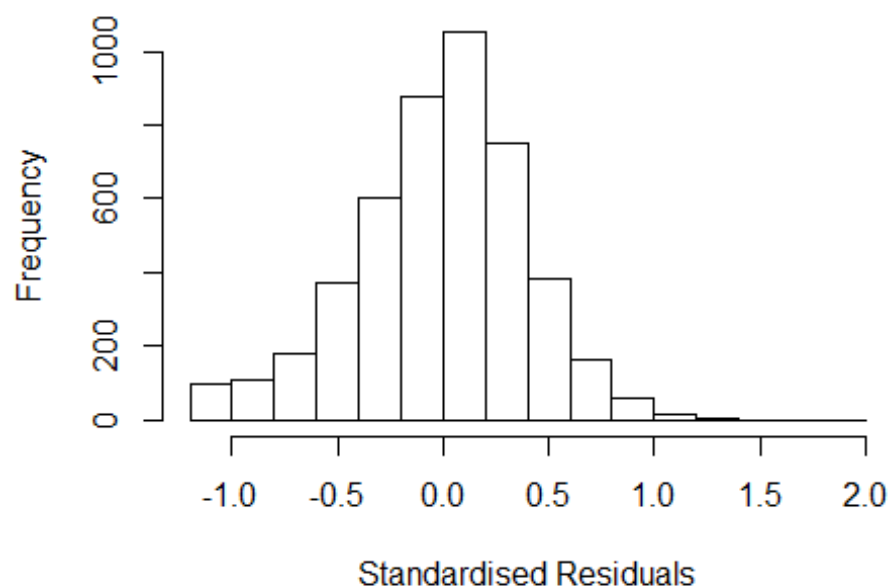
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2708 -0.2513  0.0163  0.2421  2.1052
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04595    0.04030   25.96 < 2e-16 ***
## FirstAuthorFemale1 -0.01550    0.01251   -1.24  0.215
## LastAuthorFemale1 -0.00558    0.01419   -0.39  0.694
## UniqueAuthors2     0.15544    0.01920    8.10 7.2e-16 ***
## UniqueAuthors3     0.16941    0.01953    8.67 < 2e-16 ***
## UniqueAuthors4     0.21331    0.02189    9.75 < 2e-16 ***
## UniqueAuthors5     0.27863    0.02310   12.06 < 2e-16 ***
## Year1997          -0.00950    0.04894   -0.19  0.846
## Year1998          -0.05298    0.05153   -1.03  0.304
## Year1999          -0.10714    0.05014   -2.14  0.033 *
```

```

## Year2000      -0.06703      0.04970      -1.35      0.178
## Year2001      -0.09310      0.04887      -1.91      0.057 .
## Year2002      -0.01909      0.04659      -0.41      0.682
## Year2003      -0.11135      0.04769      -2.33      0.020 *
## Year2004      -0.06272      0.04690      -1.34      0.181
## Year2005      -0.07293      0.04461      -1.63      0.102
## Year2006      -0.06829      0.04421      -1.54      0.122
## Year2007      -0.08218      0.04428      -1.86      0.064 .
## Year2008      -0.06481      0.04355      -1.49      0.137
## Year2009      -0.06010      0.04489      -1.34      0.181
## Year2010      -0.07691      0.04352      -1.77      0.077 .
## Year2011      -0.04518      0.04442      -1.02      0.309
## Year2012      -0.05376      0.04453      -1.21      0.227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0444
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 922 is an outlier with |weight| = 0 ( < 2.1e-05);
## 400 weights are ~= 1. The remaining 4270 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0238 0.8590 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           2.14e-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 1.075 1 1.037
## LastAuthorFemale 1.056 1 1.027
## Year 1.057 16 1.002

```

Residuals from first and last author



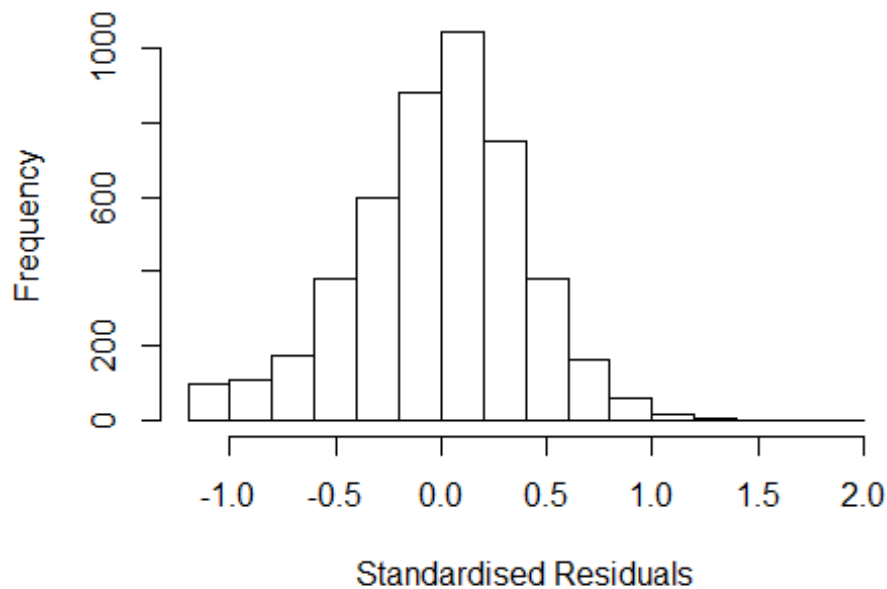
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.178 -0.256 0.015 0.250 1.964
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.151084 0.039916 28.84 <2e-16 ***
## FirstAuthorFemale1 -0.007677 0.012659 -0.61 0.544
## LastAuthorFemale1 -0.013194 0.014404 -0.92 0.360
## Year1997 -0.021082 0.049796 -0.42 0.672
## Year1998 -0.059977 0.052568 -1.14 0.254
## Year1999 -0.085726 0.051296 -1.67 0.095 .
## Year2000 -0.043745 0.051320 -0.85 0.394
## Year2001 -0.056974 0.049369 -1.15 0.249
## Year2002 0.000842 0.047453 0.02 0.986
## Year2003 -0.074428 0.048872 -1.52 0.128
## Year2004 -0.032129 0.048087 -0.67 0.504
## Year2005 -0.033306 0.045494 -0.73 0.464
```

```

## Year2006      -0.014581    0.044995   -0.32    0.746
## Year2007      -0.018285    0.044701   -0.41    0.683
## Year2008      -0.010085    0.044256   -0.23    0.820
## Year2009       0.003569    0.045645    0.08    0.938
## Year2010      -0.003734    0.044013   -0.08    0.932
## Year2011       0.021423    0.045285    0.47    0.636
## Year2012       0.026857    0.045126    0.60    0.552
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.00603,    Adjusted R-squared:  0.00219
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 922 is an outlier with |weight| = 0 ( < 2.1e-05);
## 431 weights are ~ = 1. The remaining 4239 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0334 0.8600 0.9480 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      2.14e-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 1.038 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
## -1.1750 -0.2549  0.0148  0.2512  1.9546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14907    0.03984   28.84  <2e-16 ***
## FirstAuthorFemale1 -0.01015    0.01245   -0.82   0.415
## Year1997        -0.02182    0.04978   -0.44   0.661
## Year1998        -0.05941    0.05253   -1.13   0.258
## Year1999        -0.08506    0.05132   -1.66   0.098 .
## Year2000        -0.04394    0.05134   -0.86   0.392
## Year2001        -0.05649    0.04938   -1.14   0.253
## Year2002         0.00122    0.04743    0.03   0.980
## Year2003        -0.07435    0.04887   -1.52   0.128
## Year2004        -0.03201    0.04807   -0.67   0.505
## Year2005        -0.03366    0.04550   -0.74   0.460
## Year2006        -0.01494    0.04500   -0.33   0.740
```

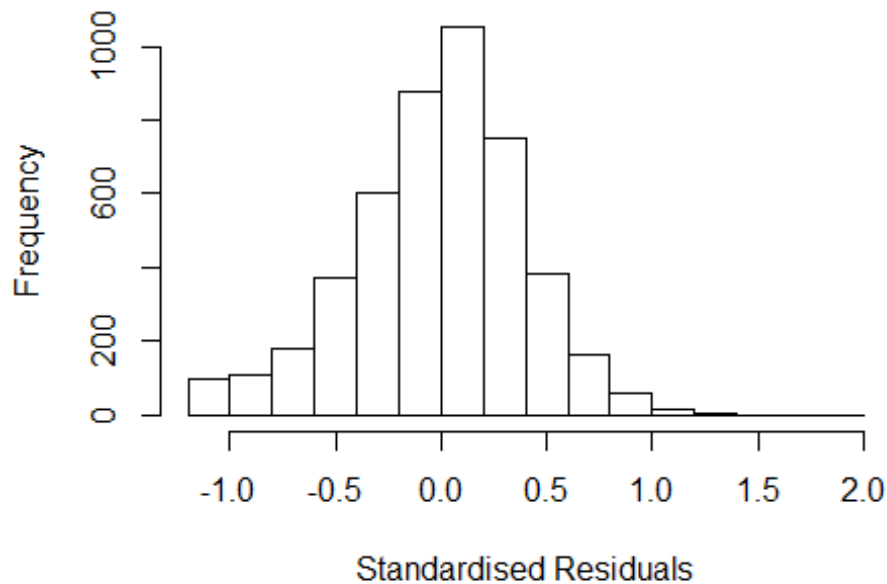


```

## Year2007          -0.01844    0.04472   -0.41    0.680
## Year2008          -0.01029    0.04427   -0.23    0.816
## Year2009           0.00371    0.04564    0.08    0.935
## Year2010          -0.00413    0.04401   -0.09    0.925
## Year2011           0.02132    0.04529    0.47    0.638
## Year2012           0.02598    0.04512    0.58    0.565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.00585,    Adjusted R-squared:  0.00222
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 922 is an outlier with |weight| = 0 ( < 2.1e-05);
## 424 weights are ~= 1. The remaining 4246 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0319 0.8610 0.9480 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      2.14e-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 1.019 1          1.010
## Year              1.019 16          1.001

```

Residuals from last author



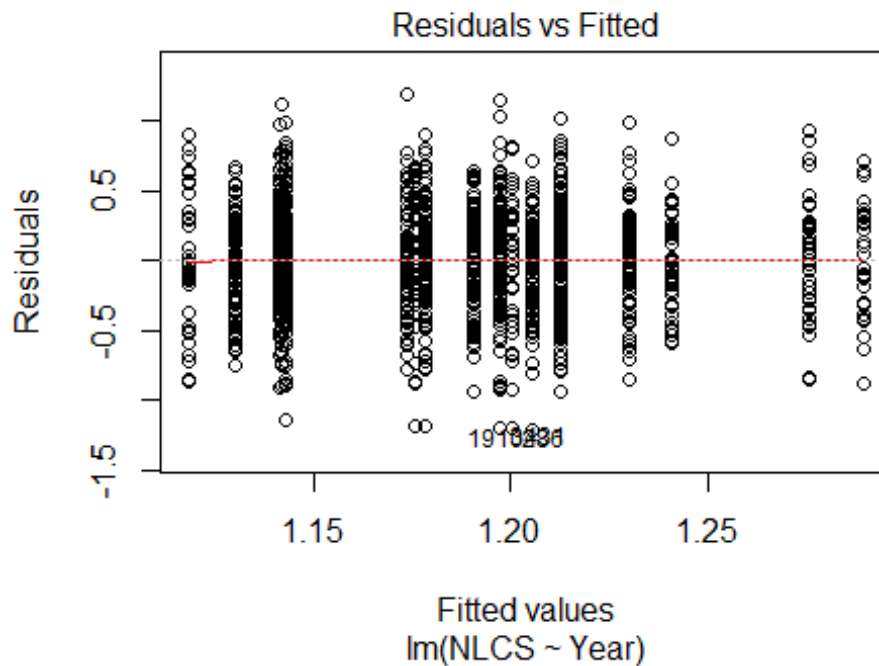
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1757 -0.2548 0.0148 0.2501 1.9600
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.150141 0.039930 28.80 <2e-16 ***
## LastAuthorFemale1 -0.015015 0.014170 -1.06 0.289
## Year1997 -0.021502 0.049791 -0.43 0.666
## Year1998 -0.060196 0.052559 -1.15 0.252
## Year1999 -0.087025 0.051148 -1.70 0.089 .
## Year2000 -0.044114 0.051332 -0.86 0.390
## Year2001 -0.058104 0.049306 -1.18 0.239
## Year2002 0.000258 0.047428 0.01 0.996
## Year2003 -0.075577 0.048753 -1.55 0.121
## Year2004 -0.033147 0.048050 -0.69 0.490
## Year2005 -0.034575 0.045389 -0.76 0.446
## Year2006 -0.015850 0.044934 -0.35 0.724
```

```

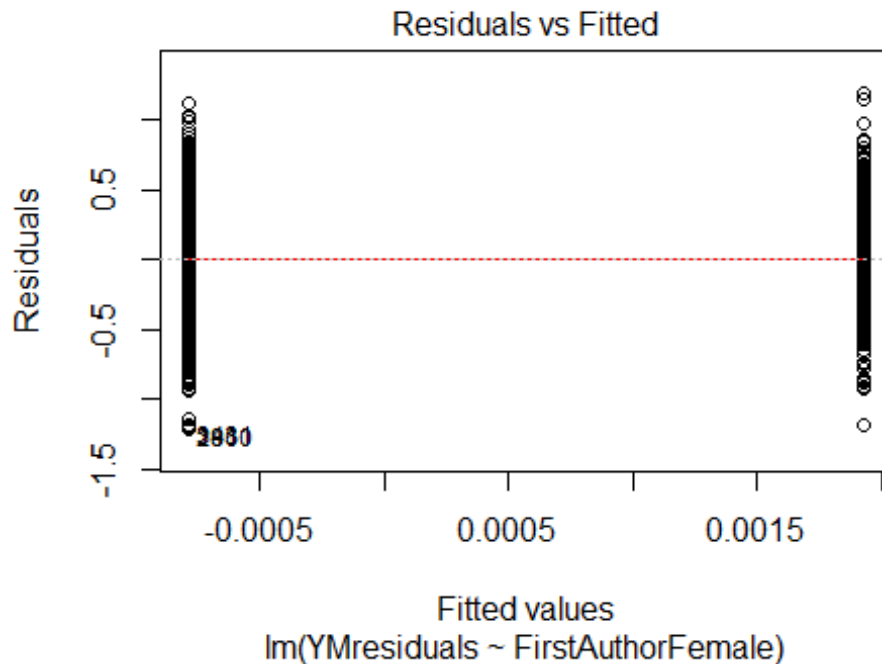
## Year2007      -0.019665    0.044602    -0.44    0.659
## Year2008      -0.011456    0.044158    -0.26    0.795
## Year2009       0.002094    0.045522     0.05    0.963
## Year2010      -0.005322    0.043907    -0.12    0.904
## Year2011       0.019377    0.045029     0.43    0.667
## Year2012       0.025544    0.045058     0.57    0.571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.00596,    Adjusted R-squared:  0.00233
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 922 is an outlier with |weight| = 0 ( < 2.1e-05);
## 421 weights are ~= 1. The remaining 4249 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0356 0.8600 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      2.14e-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: 4671"
## [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
## 112 136 128 94 141 132 157 146 170 171 188 159 180 221 223
## 2011 2012
## 260 242
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 37 41 39 50 35 76 71 95 100 101 94 103 135 127

```

```
## 2011 2012
## 157 138
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 30 29 35 42 31 57 57 84 84 82 84 87 114 103
## 2011 2012
## 128 119
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 42, df = 16, p-value = 4e-04
```

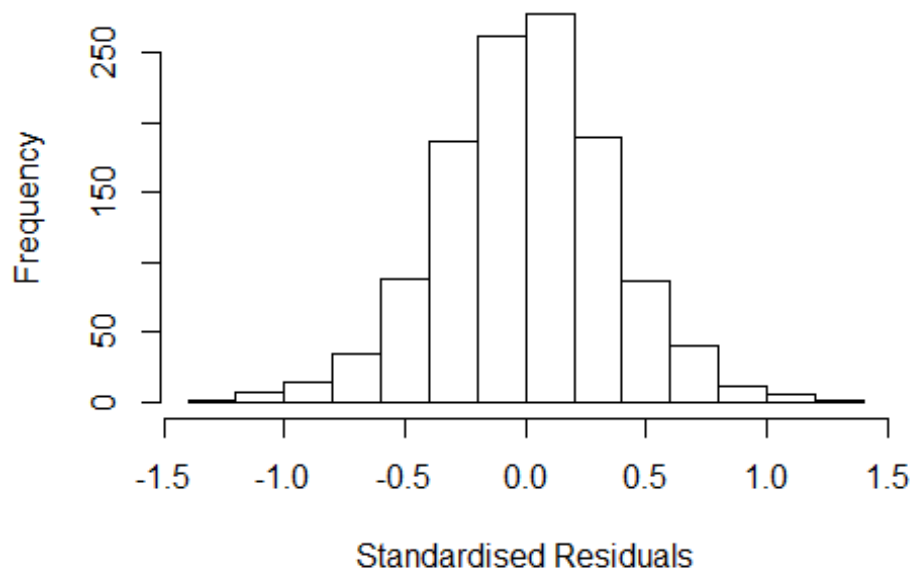


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.38664409986715"
## [1] "Male first author team size 2018 geometric mean: 3.88401434169778"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.08473395163359"
## [1] "Male last author team size 2018 geometric mean: 4.06292700665372"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, 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 1.120 1      1.059
## LastAuthorFemale  1.080 1      1.039
## UniqueAuthors     1.472 4      1.050
## Year              1.606 16     1.015
```

Residuals from first and last author and team size



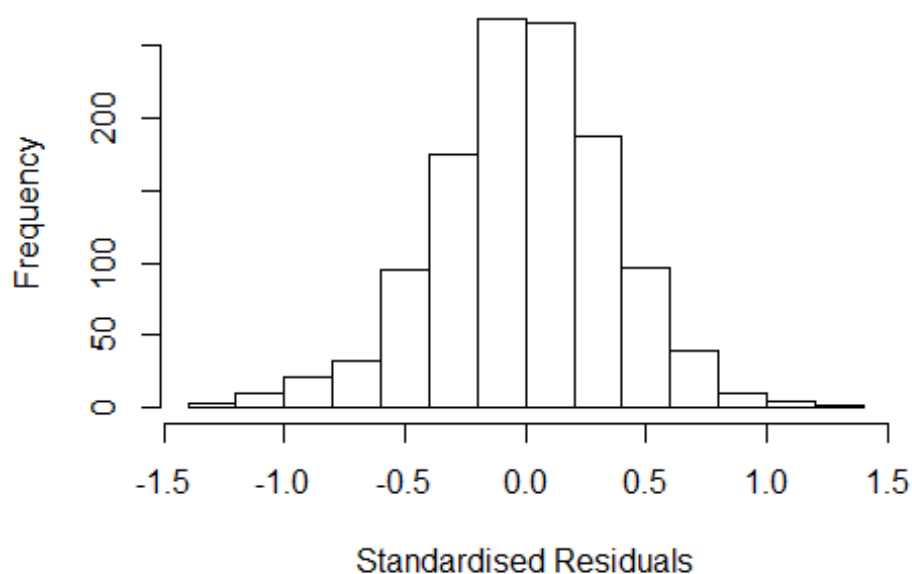
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30245 -0.22746  0.00558  0.21986  1.22506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.079963   0.081432   13.26 < 2e-16 ***
## FirstAuthorFemale1 -0.015861   0.022658   -0.70  0.4841
## LastAuthorFemale1  0.079256   0.025349    3.13  0.0018 **
## UniqueAuthors2     0.222484   0.044384    5.01 6.2e-07 ***
## UniqueAuthors3     0.175981   0.044579    3.95 8.4e-05 ***
## UniqueAuthors4     0.158963   0.046222    3.44 0.0006 ***
## UniqueAuthors5     0.277807   0.044062    6.30 4.1e-10 ***
## Year1997           0.040198   0.117172    0.34  0.7316
## Year1998           0.151202   0.116764    1.29  0.1956
## Year1999          -0.130484   0.132516   -0.98  0.3250
```

```

## Year2000      -0.022071    0.084053   -0.26    0.7929
## Year2001      0.000369    0.100404    0.00    0.9971
## Year2002     -0.111026    0.083728   -1.33    0.1851
## Year2003     -0.053165    0.085175   -0.62    0.5326
## Year2004     -0.138605    0.081174   -1.71    0.0880 .
## Year2005     -0.158638    0.079507   -2.00    0.0462 *
## Year2006     -0.069351    0.077307   -0.90    0.3699
## Year2007     -0.122035    0.080241   -1.52    0.1286
## Year2008     -0.069572    0.079121   -0.88    0.3794
## Year2009     -0.138715    0.079408   -1.75    0.0809 .
## Year2010     -0.100034    0.076821   -1.30    0.1931
## Year2011     -0.119517    0.075779   -1.58    0.1150
## Year2012     -0.119147    0.078478   -1.52    0.1292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0713, Adjusted R-squared:  0.0541
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1097 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.103  0.862  0.952   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          8.28e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.073 1          1.036
## LastAuthorFemale  1.065 1          1.032
## 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
## -1.290016 -0.240642 -0.000789  0.234182  1.228546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.24554    0.06883   18.10  <2e-16 ***
## FirstAuthorFemale1 -0.00858    0.02273   -0.38   0.7058
## LastAuthorFemale1  0.07353    0.02591    2.84   0.0046 **
## Year1997          0.04447    0.11525    0.39   0.6997
## Year1998          0.10182    0.11255    0.90   0.3658
## Year1999         -0.13369    0.13360   -1.00   0.3172
## Year2000         -0.03301    0.08319   -0.40   0.6916
## Year2001         -0.00752    0.10448   -0.07   0.9426
## Year2002         -0.11231    0.08363   -1.34   0.1795
## Year2003         -0.03693    0.08576   -0.43   0.6668
## Year2004         -0.12733    0.08020   -1.59   0.1126
## Year2005         -0.14627    0.07843   -1.86   0.0624 .
```

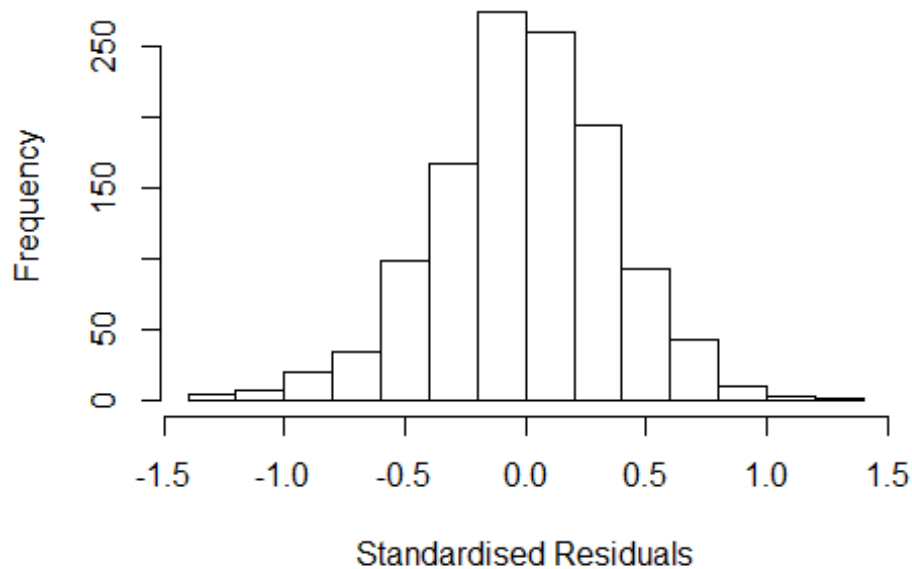


```

## Year2006          -0.04659    0.07640   -0.61    0.5421
## Year2007          -0.10051    0.07928   -1.27    0.2051
## Year2008          -0.05128    0.07698   -0.67    0.5054
## Year2009          -0.10917    0.07944   -1.37    0.1696
## Year2010          -0.06361    0.07652   -0.83    0.4060
## Year2011          -0.07877    0.07575   -1.04    0.2986
## Year2012          -0.06946    0.07716   -0.90    0.3682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.00943
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 103 weights are ~= 1. The remaining 1105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.145  0.869  0.950  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      8.28e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"    "nonsingular"      ".vcov.avar1"
## 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



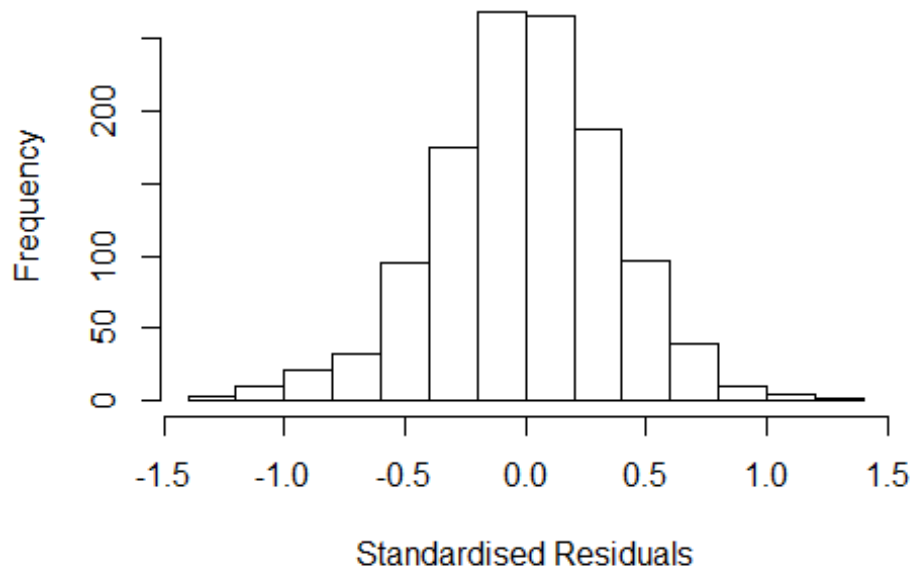
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.290467 -0.240692 -0.000549 0.234250 1.205377
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.249882 0.069872 17.89 <2e-16 ***
## FirstAuthorFemale1 0.001250 0.022958 0.05 0.957
## Year1997 0.040584 0.115512 0.35 0.725
## Year1998 0.100989 0.113830 0.89 0.375
## Year1999 -0.132244 0.130244 -1.02 0.310
## Year2000 -0.029977 0.084328 -0.36 0.722
## Year2001 -0.000909 0.104979 -0.01 0.993
## Year2002 -0.106342 0.084481 -1.26 0.208
## Year2003 -0.034395 0.086765 -0.40 0.692
## Year2004 -0.124192 0.081191 -1.53 0.126
## Year2005 -0.135523 0.078949 -1.72 0.086 .
## Year2006 -0.037378 0.077363 -0.48 0.629
```

```

## Year2007          -0.091510    0.080121   -1.14    0.254
## Year2008          -0.038974    0.077757   -0.50    0.616
## Year2009          -0.100191    0.080086   -1.25    0.211
## Year2010          -0.054071    0.077166   -0.70    0.484
## Year2011          -0.072331    0.076603   -0.94    0.345
## Year2012          -0.058132    0.077818   -0.75    0.455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.352
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.00367
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1106 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.870  0.951  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      8.28e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



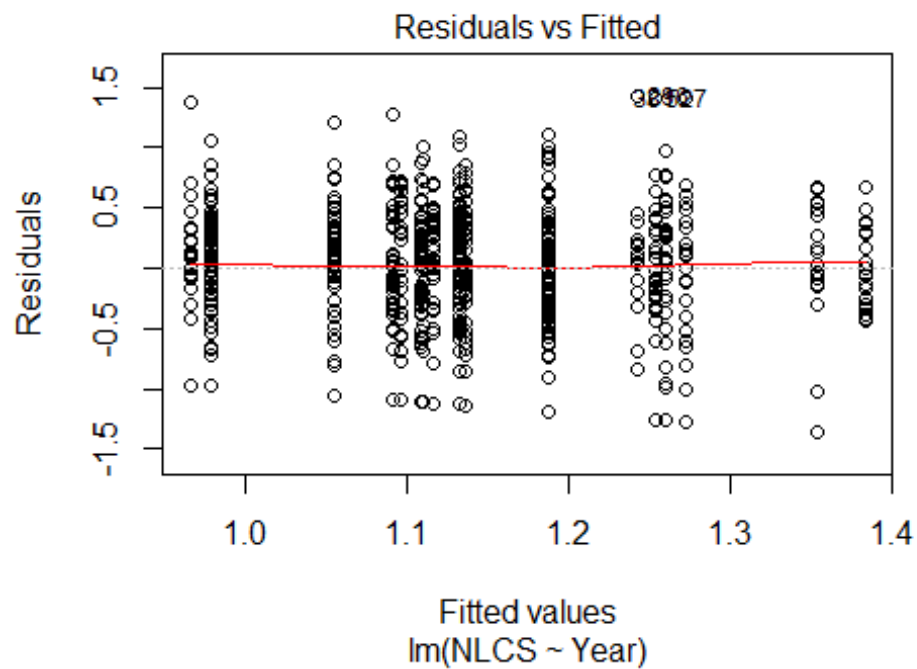
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28739 -0.24095 -0.00247 0.23560 1.22187
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2438 0.0688 18.08 <2e-16 ***
## LastAuthorFemale1 0.0720 0.0260 2.77 0.0058 **
## Year1997 0.0436 0.1150 0.38 0.7048
## Year1998 0.1017 0.1126 0.90 0.3666
## Year1999 -0.1351 0.1327 -1.02 0.3090
## Year2000 -0.0331 0.0832 -0.40 0.6903
## Year2001 -0.0091 0.1033 -0.09 0.9299
## Year2002 -0.1124 0.0836 -1.34 0.1792
## Year2003 -0.0386 0.0853 -0.45 0.6506
## Year2004 -0.1286 0.0799 -1.61 0.1077
## Year2005 -0.1472 0.0783 -1.88 0.0603 .
## Year2006 -0.0470 0.0764 -0.62 0.5384
```

```

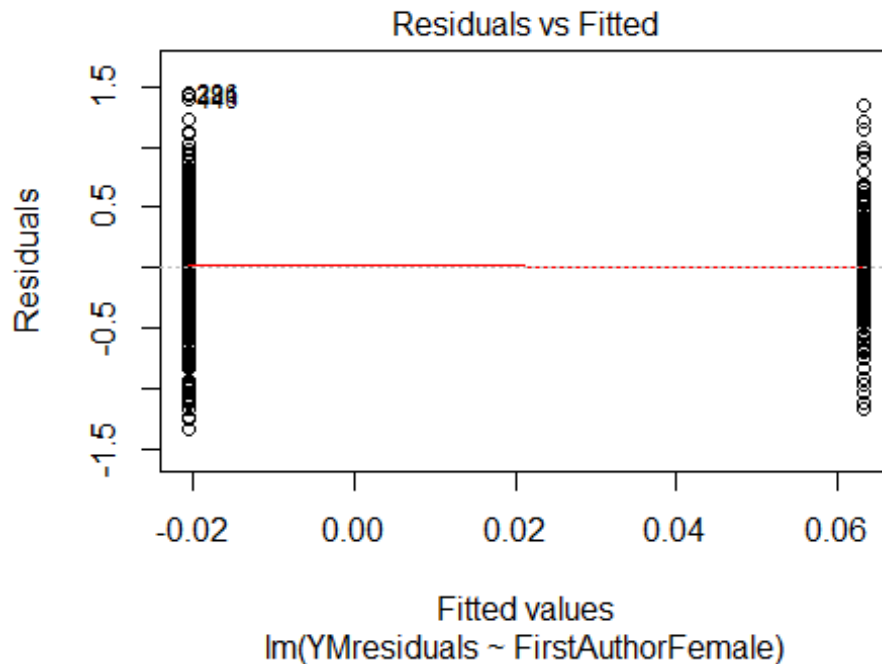
## Year2007          -0.1007      0.0793   -1.27    0.2042
## Year2008          -0.0524      0.0768   -0.68    0.4948
## Year2009          -0.1099      0.0792   -1.39    0.1655
## Year2010          -0.0643      0.0763   -0.84    0.3996
## Year2011          -0.0791      0.0757   -1.04    0.2965
## Year2012          -0.0701      0.0771   -0.91    0.3633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.351
## Multiple R-squared:  0.0241, Adjusted R-squared:  0.0101
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1106 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.149  0.870  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      8.28e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1208"
## [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
##   47   59   58   49   54   70   54   59   68   71  105   94   97  172  167
## 2011 2012
##  209  201
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   20   29   24   15   17   25   25   29   30   34   50   50   62   94   63
## 2011 2012

```

```
## 102 101
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 27 22 14 15 23 23 28 25 33 42 46 56 77 59
## 2011 2012
## 81 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 = 24, df = 16, p-value = 0.09
```

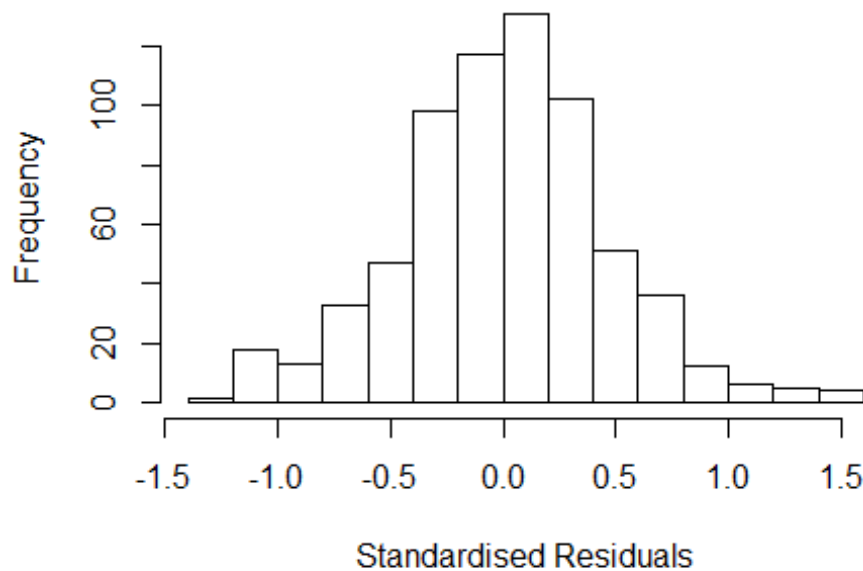


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.2, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.18349167992907"
## [1] "Male first author team size 2018 geometric mean: 3.49011416470914"
##
## 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: 4.15917458405027"
## [1] "Male last author team size 2018 geometric mean: 3.55681266124639"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.132 1      1.064
## LastAuthorFemale  1.265 1      1.125
## UniqueAuthors     1.922 4      1.085
## Year              2.433 16      1.028
```

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
## -1.3791 -0.2873 0.0108 0.2795 1.5562
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0617 0.1700 6.25 7.6e-10 ***
## FirstAuthorFemale1 0.0905 0.0387 2.34 0.020 *
## LastAuthorFemale1 0.0218 0.0468 0.47 0.642
## UniqueAuthors2 0.3174 0.0692 4.59 5.4e-06 ***
## UniqueAuthors3 0.3491 0.0678 5.15 3.5e-07 ***
## UniqueAuthors4 0.3203 0.0760 4.21 2.9e-05 ***
## UniqueAuthors5 0.3962 0.0731 5.42 8.5e-08 ***
## Year1997 -0.0448 0.1963 -0.23 0.820
## Year1998 0.1013 0.1782 0.57 0.570
## Year1999 -0.2524 0.2126 -1.19 0.236
```

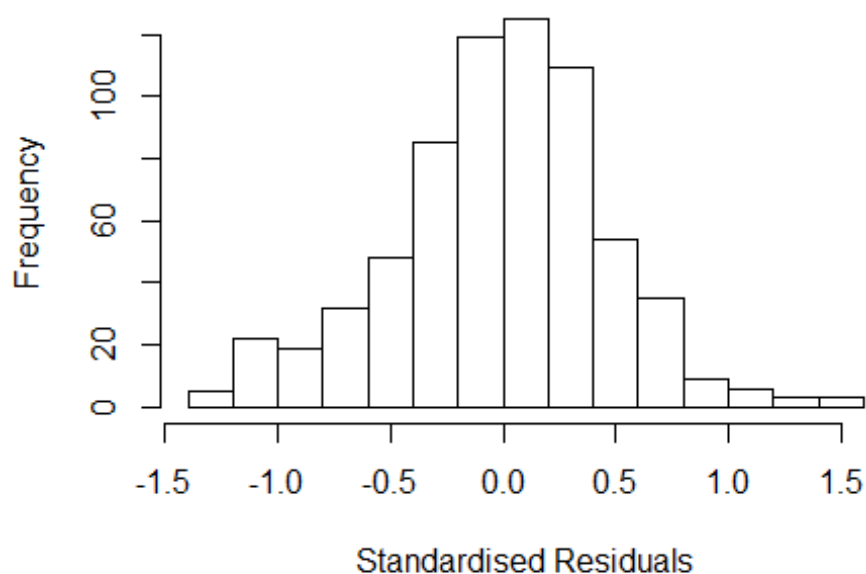


```

## Year2000          -0.1459      0.2102   -0.69    0.488
## Year2001          -0.0411      0.1960   -0.21    0.834
## Year2002          -0.2729      0.1857   -1.47    0.142
## Year2003          -0.2332      0.1734   -1.35    0.179
## Year2004          -0.3124      0.1854   -1.69    0.092 .
## Year2005          -0.2342      0.1777   -1.32    0.188
## Year2006          -0.2717      0.1709   -1.59    0.112
## Year2007          -0.2715      0.1712   -1.59    0.113
## Year2008          -0.3093      0.1727   -1.79    0.074 .
## Year2009          -0.3673      0.1647   -2.23    0.026 *
## Year2010          -0.2844      0.1654   -1.72    0.086 .
## Year2011          -0.2473      0.1629   -1.52    0.129
## Year2012          -0.2529      0.1642   -1.54    0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.13,   Adjusted R-squared:  0.101
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 610 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.147  0.857  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           1.48e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.128 1           1.062
## LastAuthorFemale  1.287 1           1.135
## Year              1.384 16           1.010

```

Residuals from first and last author



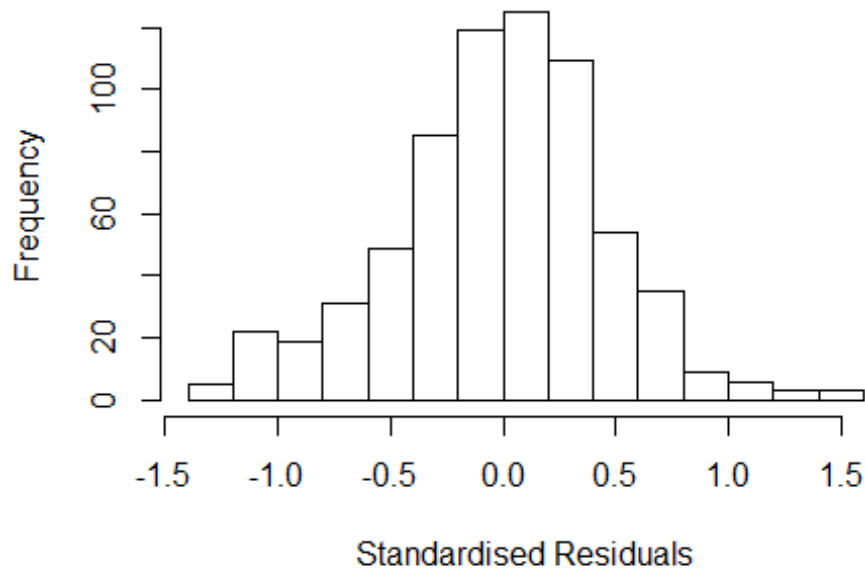
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32921 -0.28837 0.00802 0.29723 1.52341
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.329215 0.134075 9.91 <2e-16 ***
## FirstAuthorFemale1 0.108971 0.039917 2.73 0.0065 **
## LastAuthorFemale1 -0.000758 0.049741 -0.02 0.9878
## Year1997 -0.071519 0.195348 -0.37 0.7144
## Year1998 0.043414 0.153079 0.28 0.7768
## Year1999 -0.335422 0.202212 -1.66 0.0976 .
## Year2000 -0.185629 0.185818 -1.00 0.3182
## Year2001 -0.099622 0.185933 -0.54 0.5923
## Year2002 -0.357811 0.165494 -2.16 0.0310 *
## Year2003 -0.170252 0.155787 -1.09 0.2749
## Year2004 -0.301550 0.181204 -1.66 0.0966 .
## Year2005 -0.219429 0.155811 -1.41 0.1595
```

```

## Year2006          -0.226969    0.152132    -1.49    0.1362
## Year2007          -0.290506    0.151124    -1.92    0.0550 .
## Year2008          -0.293343    0.154359    -1.90    0.0578 .
## Year2009          -0.373210    0.145152    -2.57    0.0104 *
## Year2010          -0.213205    0.144569    -1.47    0.1408
## Year2011          -0.198166    0.141965    -1.40    0.1632
## Year2012          -0.188794    0.142103    -1.33    0.1845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0523, Adjusted R-squared:  0.0262
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 627 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.171  0.857   0.953   0.887   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      1.48e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.109 1          1.053
## Year              1.109 16          1.003

```

Residuals from first author



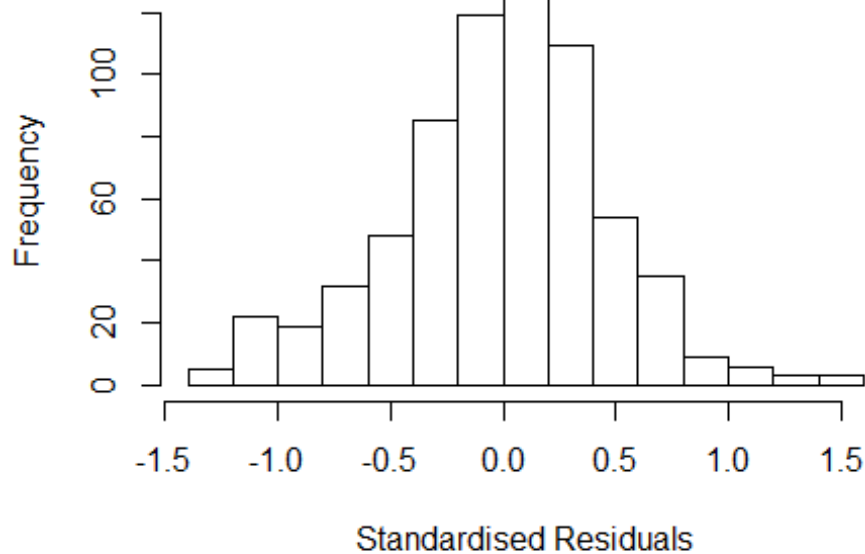
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32973 -0.28846 0.00829 0.29715 1.52360
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3297 0.1340 9.93 <2e-16 ***
## FirstAuthorFemale1 0.1087 0.0399 2.72 0.0066 **
## Year1997 -0.0718 0.1952 -0.37 0.7130
## Year1998 0.0428 0.1530 0.28 0.7795
## Year1999 -0.3362 0.2010 -1.67 0.0948 .
## Year2000 -0.1863 0.1853 -1.01 0.3151
## Year2001 -0.1002 0.1860 -0.54 0.5905
## Year2002 -0.3586 0.1631 -2.20 0.0282 *
## Year2003 -0.1708 0.1557 -1.10 0.2728
## Year2004 -0.3020 0.1815 -1.66 0.0965 .
## Year2005 -0.2199 0.1556 -1.41 0.1581
## Year2006 -0.2275 0.1520 -1.50 0.1351
```

```

## Year2007          -0.2912      0.1499   -1.94   0.0524 .
## Year2008          -0.2938      0.1545   -1.90   0.0577 .
## Year2009          -0.3738      0.1443   -2.59   0.0098 **
## Year2010          -0.2138      0.1437   -1.49   0.1373
## Year2011          -0.1988      0.1419   -1.40   0.1619
## Year2012          -0.1894      0.1415   -1.34   0.1814
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.424
## Multiple R-squared:  0.0523, Adjusted R-squared:  0.0278
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 627 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.169  0.857   0.952   0.887   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      1.48e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.25 1          1.118
## Year              1.25 16          1.007

```

Residuals from last author



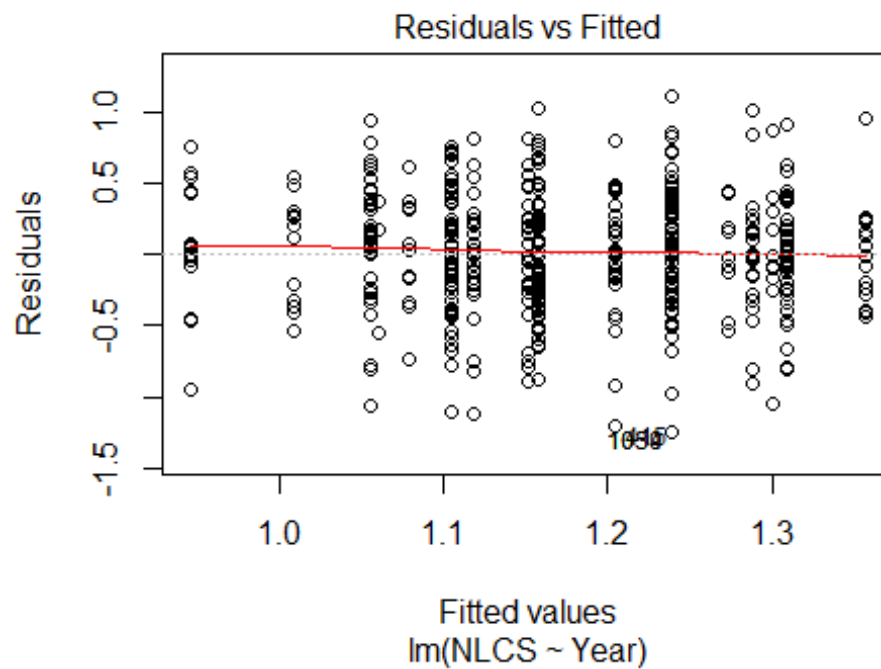
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3440 -0.2906  0.0168  0.2885  1.5064
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3440     0.1417   9.49  <2e-16 ***
## LastAuthorFemale1  0.0331     0.0488   0.68   0.497
## Year1997        -0.0510     0.2005  -0.25   0.799
## Year1998         0.0352     0.1595   0.22   0.825
## Year1999        -0.3304     0.2137  -1.55   0.123
## Year2000        -0.1834     0.1889  -0.97   0.332
## Year2001        -0.0822     0.1920  -0.43   0.669
## Year2002        -0.3519     0.1744  -2.02   0.044 *
## Year2003        -0.1367     0.1628  -0.84   0.401
## Year2004        -0.2984     0.1855  -1.61   0.108
## Year2005        -0.2092     0.1615  -1.30   0.196
## Year2006        -0.2244     0.1585  -1.42   0.157
```

```

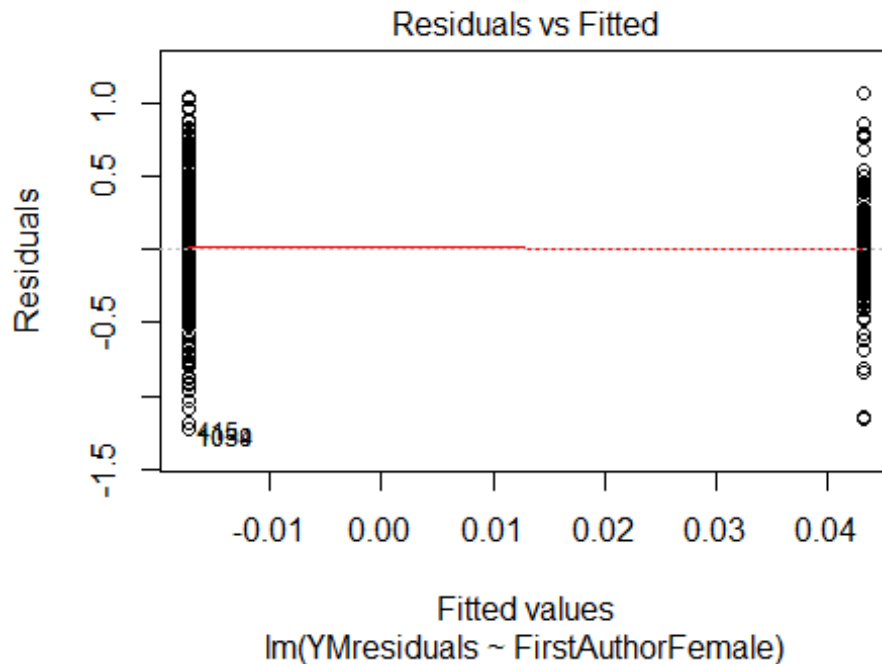
## Year2007          -0.2863      0.1571    -1.82      0.069 .
## Year2008          -0.2830      0.1604    -1.77      0.078 .
## Year2009          -0.3582      0.1515    -2.36      0.018 *
## Year2010          -0.2101      0.1523    -1.38      0.168
## Year2011          -0.1909      0.1491    -1.28      0.201
## Year2012          -0.1793      0.1496    -1.20      0.231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0431, Adjusted R-squared:  0.0183
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 624 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.183  0.860  0.951  0.886  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      1.48e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 674"
## [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
##   31   10   29   23   26   30   32   20   25   45   38   45   53   59   90
## 2011 2012
##   94   95
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15    3   11   17   11    9   22   13   17   33   29   35   42   36   65
## 2011 2012

```

```
## 62 65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 2 9 17 11 7 21 12 17 30 29 35 37 32 57
## 2011 2012
## 61 63
## [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
```

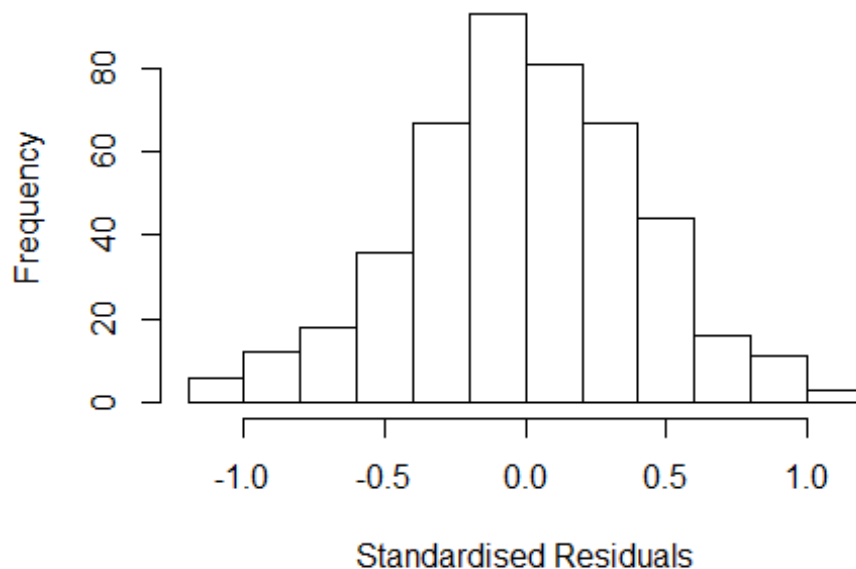


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.8, df = 1, p-value = 0.03
```

```
## [1] "Female first author team size 2018 geometric mean: 3.38246084535204"
## [1] "Male first author team size 2018 geometric mean: 3.20307227777185"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.9617501952031"
## [1] "Male last author team size 2018 geometric mean: 3.40759256958854"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 800, 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 1.374 1      1.172
## LastAuthorFemale  1.288 1      1.135
## UniqueAuthors    2.213 4      1.104
## Year             2.806 16      1.033
```

Residuals from first and last author and team size



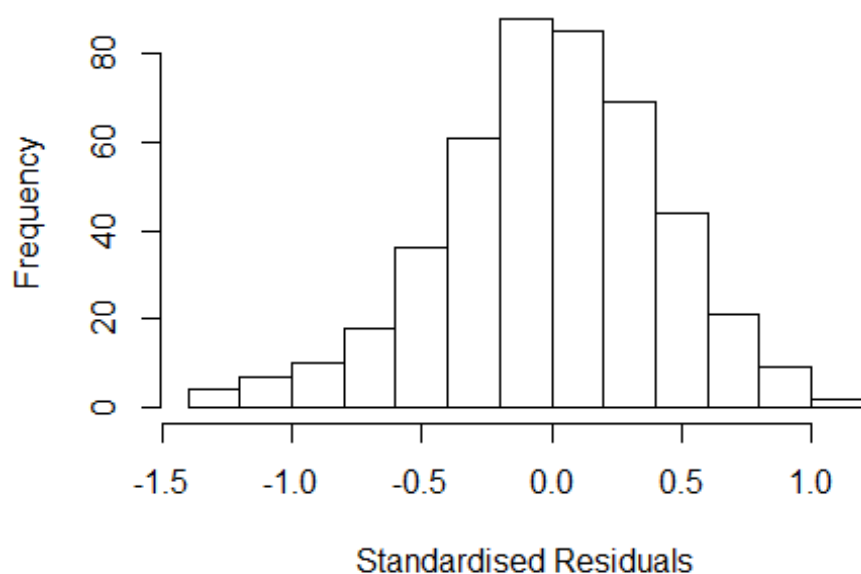
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.07173 -0.26683 -0.00595 0.27045 1.15793
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23383 0.10453 11.80 < 2e-16 ***
## FirstAuthorFemale1 0.01871 0.04427 0.42 0.67282
## LastAuthorFemale1 0.10972 0.04422 2.48 0.01347 *
## UniqueAuthors2 0.15505 0.06111 2.54 0.01153 *
## UniqueAuthors3 0.17029 0.06624 2.57 0.01048 *
## UniqueAuthors4 0.27701 0.07820 3.54 0.00044 ***
## UniqueAuthors5 0.24573 0.06913 3.55 0.00042 ***
## Year1997 -0.44048 0.41786 -1.05 0.29242
## Year1998 0.00838 0.18942 0.04 0.96472
## Year1999 -0.43138 0.15892 -2.71 0.00691 **
```

```

## Year2000      -0.33626    0.14035   -2.40  0.01700 *
## Year2001      -0.13566    0.17984   -0.75  0.45107
## Year2002      -0.09077    0.13586   -0.67  0.50442
## Year2003      -0.37724    0.14682   -2.57  0.01052 *
## Year2004      -0.25175    0.14494   -1.74  0.08311 .
## Year2005      -0.29053    0.12815   -2.27  0.02387 *
## Year2006      -0.24691    0.13229   -1.87  0.06266 .
## Year2007      -0.21607    0.12667   -1.71  0.08877 .
## Year2008      -0.31845    0.14050   -2.27  0.02392 *
## Year2009      -0.11965    0.13333   -0.90  0.37002
## Year2010      -0.30026    0.12211   -2.46  0.01433 *
## Year2011      -0.27647    0.11954   -2.31  0.02121 *
## Year2012      -0.20079    0.12087   -1.66  0.09740 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0629
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 411 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.419  0.880  0.952  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          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 1.321 1          1.149
## LastAuthorFemale 1.193 1          1.092
## Year 1.388 16          1.010

```

Residuals from first and last author



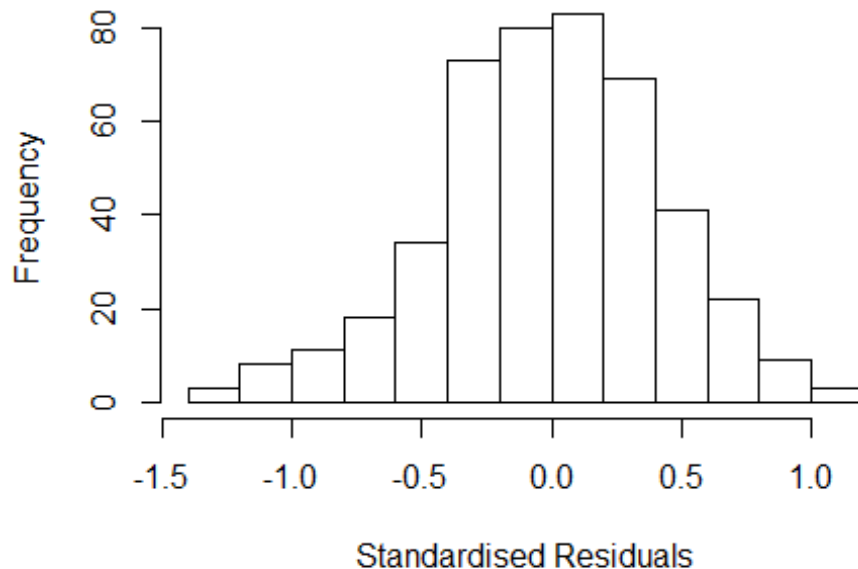
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2307 -0.2529 0.0037 0.2729 1.1163
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3110 0.0924 14.19 <2e-16 ***
## FirstAuthorFemale1 0.0234 0.0451 0.52 0.6042
## LastAuthorFemale1 0.1191 0.0449 2.65 0.0084 **
## Year1997 -0.4325 0.3115 -1.39 0.1657
## Year1998 0.0269 0.1753 0.15 0.8780
## Year1999 -0.3577 0.1517 -2.36 0.0188 *
## Year2000 -0.3087 0.1544 -2.00 0.0462 *
## Year2001 -0.0544 0.1758 -0.31 0.7570
## Year2002 -0.0501 0.1277 -0.39 0.6949
## Year2003 -0.2821 0.1436 -1.96 0.0501 .
## Year2004 -0.2373 0.1364 -1.74 0.0827 .
## Year2005 -0.2377 0.1165 -2.04 0.0419 *
```

```

## Year2006          -0.1787      0.1262    -1.42    0.1573
## Year2007          -0.1285      0.1105    -1.16    0.2456
## Year2008          -0.2591      0.1326    -1.95    0.0514 .
## Year2009          -0.0500      0.1175    -0.43    0.6704
## Year2010          -0.2229      0.1109    -2.01    0.0450 *
## Year2011          -0.2010      0.1076    -1.87    0.0625 .
## Year2012          -0.1047      0.1085    -0.96    0.3351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.415
## Multiple R-squared:  0.0574, Adjusted R-squared:  0.0184
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 409 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.359  0.879  0.954  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      2.20e-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 1.245 1          1.116
## Year              1.245 16          1.007

```

Residuals from first author



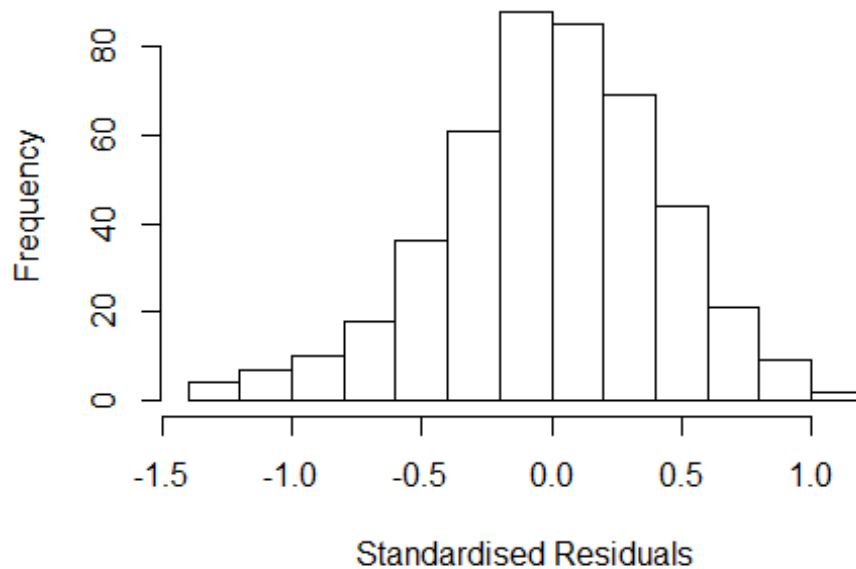
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.229511 -0.260586 -0.000319 0.272486 1.061143
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3286 0.0914 14.53 <2e-16 ***
## FirstAuthorFemale1 0.0553 0.0439 1.26 0.208
## Year1997 -0.4501 0.3096 -1.45 0.147
## Year1998 0.0263 0.1794 0.15 0.884
## Year1999 -0.3757 0.1531 -2.45 0.015 *
## Year2000 -0.2588 0.1568 -1.65 0.100 .
## Year2001 -0.0804 0.1701 -0.47 0.637
## Year2002 -0.0597 0.1290 -0.46 0.644
## Year2003 -0.2927 0.1452 -2.02 0.044 *
## Year2004 -0.2358 0.1363 -1.73 0.084 .
## Year2005 -0.2167 0.1182 -1.83 0.067 .
## Year2006 -0.1734 0.1251 -1.39 0.167
```

```

## Year2007          -0.1173      0.1093   -1.07    0.284
## Year2008          -0.2510      0.1328   -1.89    0.059 .
## Year2009          -0.0330      0.1162   -0.28    0.776
## Year2010          -0.2209      0.1105   -2.00    0.046 *
## Year2011          -0.2064      0.1074   -1.92    0.055 .
## Year2012          -0.0991      0.1083   -0.91    0.361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0453, Adjusted R-squared:  0.00804
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 410 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.376  0.886  0.952  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      2.20e-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 1.122 1          1.059
## Year            1.122 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
## -1.22099 -0.25658 0.00951 0.27016 1.13485
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3118 0.0924 14.20 <2e-16 ***
## LastAuthorFemale1 0.1257 0.0434 2.89 0.004 **
## Year1997 -0.4333 0.3118 -1.39 0.165
## Year1998 0.0253 0.1749 0.14 0.885
## Year1999 -0.3481 0.1481 -2.35 0.019 *
## Year2000 -0.3072 0.1523 -2.02 0.044 *
## Year2001 -0.0493 0.1798 -0.27 0.784
## Year2002 -0.0458 0.1281 -0.36 0.721
## Year2003 -0.2814 0.1441 -1.95 0.052 .
## Year2004 -0.2268 0.1342 -1.69 0.092 .
## Year2005 -0.2318 0.1157 -2.00 0.046 *
## Year2006 -0.1763 0.1261 -1.40 0.163
```

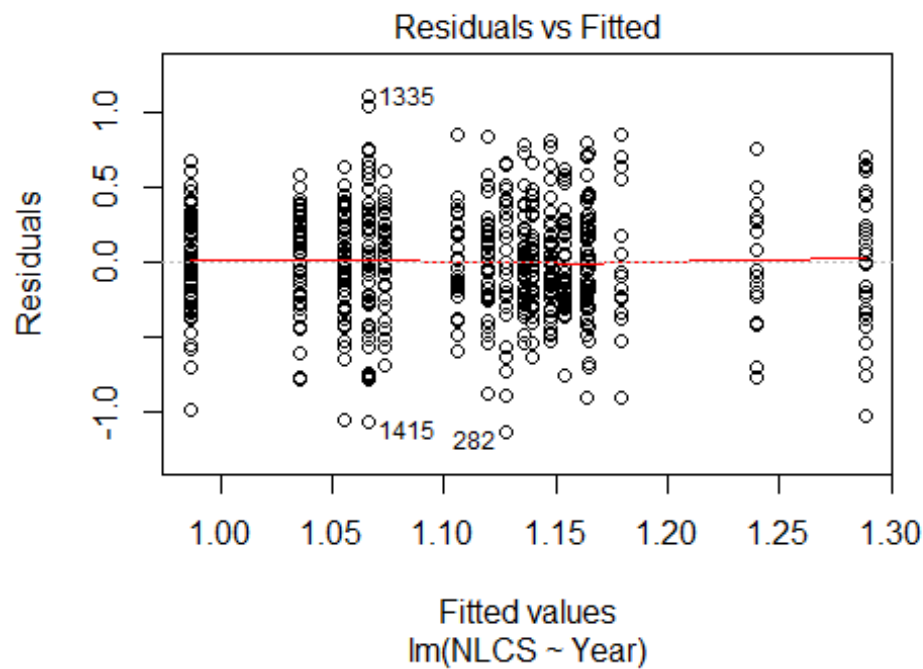


```

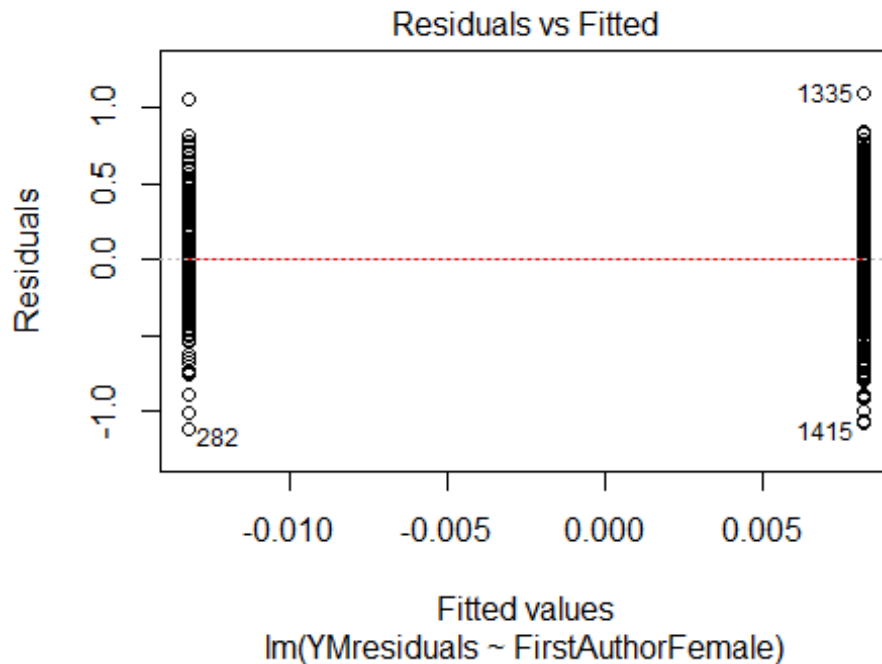
## Year2007          -0.1235      0.1104    -1.12      0.264
## Year2008          -0.2537      0.1311    -1.93      0.054 .
## Year2009          -0.0491      0.1176    -0.42      0.677
## Year2010          -0.2166      0.1102    -1.96      0.050 .
## Year2011          -0.1934      0.1063    -1.82      0.070 .
## Year2012          -0.1007      0.1086    -0.93      0.354
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0565, Adjusted R-squared:  0.0198
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 414 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.363  0.879  0.956  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      2.20e-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 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
##   30   37   51   56   63   82   54   67   57   57   70   51   75   83  100
## 2011 2012
##   89   98
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   18   22   27   32   41   31   39   34   31   45   28   45   53   65
## 2011 2012

```

```
## 55 60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 14 20 25 31 38 27 35 30 25 40 22 40 41 58
## 2011 2012
## 50 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 = 36, df = 16, p-value = 0.002
```

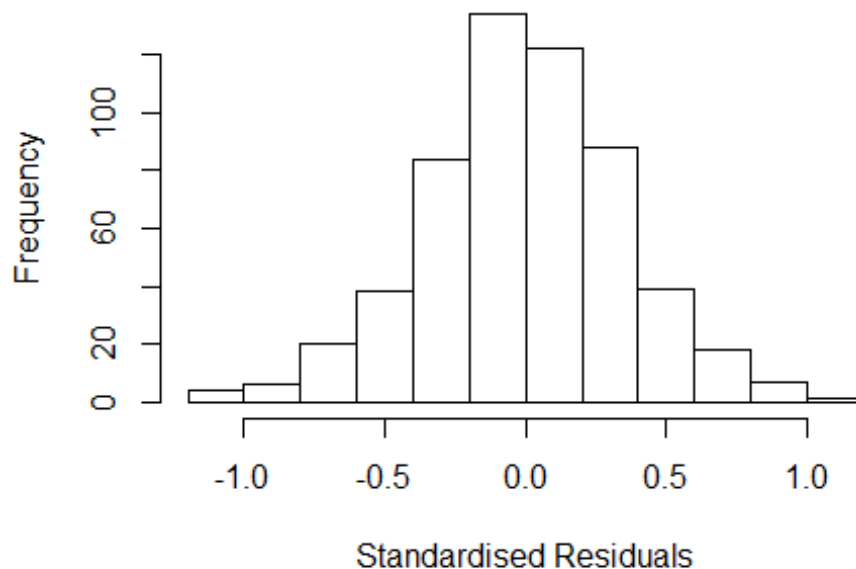


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.451112715014"
## [1] "Male first author team size 2018 geometric mean: 3.90114738364596"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.17129334639223"
## [1] "Male last author team size 2018 geometric mean: 4.26933076126292"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, 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 1.138 1      1.067
## LastAuthorFemale  1.158 1      1.076
## UniqueAuthors    1.960 4      1.088
## Year              2.398 16     1.028
```

Residuals from first and last author and team size



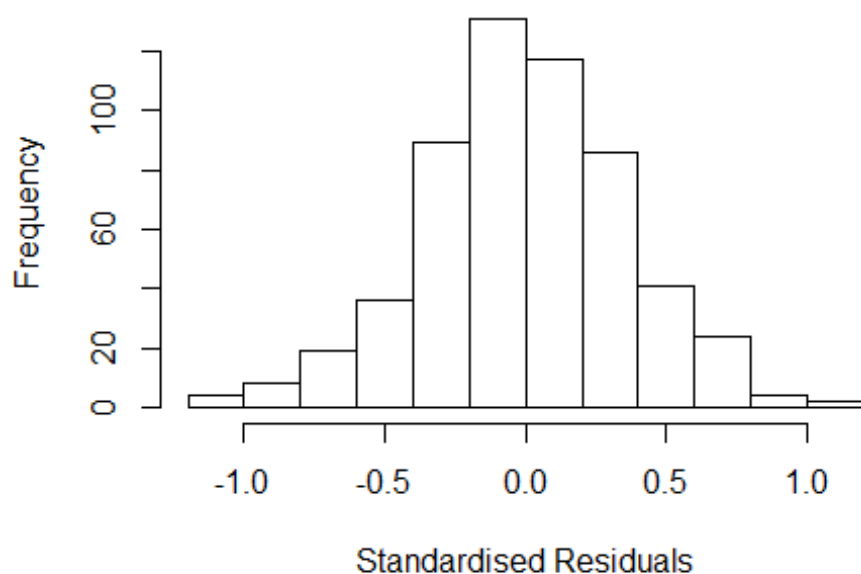
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16376 -0.21593 -0.00914  0.22716  1.03812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0892    0.1549   7.03 6.2e-12 ***
## FirstAuthorFemale1 -0.0313    0.0315  -0.99  0.3206
## LastAuthorFemale1 -0.0374    0.0373  -1.00  0.3167
## UniqueAuthors2     0.0805    0.0652   1.24  0.2171
## UniqueAuthors3     0.0216    0.0635   0.34  0.7336
## UniqueAuthors4     0.1392    0.0668   2.08  0.0376 *
## UniqueAuthors5     0.1840    0.0666   2.76  0.0059 **
## Year1997           0.1274    0.1851   0.69  0.4918
## Year1998           0.0637    0.1790   0.36  0.7221
## Year1999           0.1456    0.1858   0.78  0.4336
```

```

## Year2000          0.0253      0.1769      0.14      0.8862
## Year2001         -0.0112      0.1649     -0.07      0.9457
## Year2002          0.0239      0.1656      0.14      0.8852
## Year2003         -0.0274      0.1616     -0.17      0.8653
## Year2004         -0.0541      0.1652     -0.33      0.7434
## Year2005         -0.0626      0.1692     -0.37      0.7113
## Year2006         -0.0202      0.1599     -0.13      0.8996
## Year2007         -0.1119      0.1696     -0.66      0.5095
## Year2008         -0.0314      0.1619     -0.19      0.8463
## Year2009         -0.1343      0.1640     -0.82      0.4131
## Year2010         -0.1608      0.1606     -1.00      0.3171
## Year2011         -0.0761      0.1617     -0.47      0.6382
## Year2012         -0.0966      0.1760     -0.55      0.5834
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.334
## Multiple R-squared:  0.0818, Adjusted R-squared:  0.0443
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.199  0.864  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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.107 1      1.052
## LastAuthorFemale  1.147 1      1.071
## Year              1.260 16      1.007

```

Residuals from first and last author



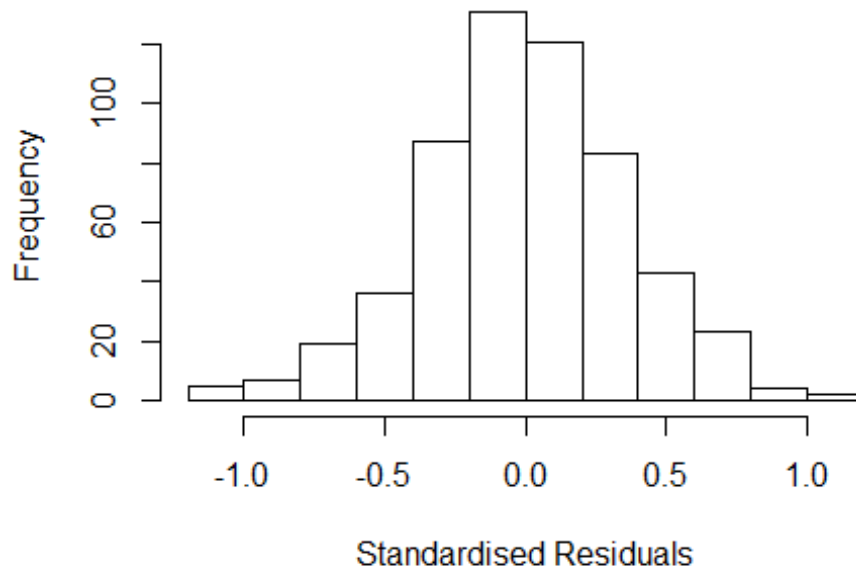
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18053 -0.22757 -0.00633 0.22209 1.07027
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15412 0.14480 7.97 9.4e-15 ***
## FirstAuthorFemale1 -0.01831 0.03190 -0.57 0.57
## LastAuthorFemale1 -0.04081 0.03710 -1.10 0.27
## Year1997 0.14686 0.17328 0.85 0.40
## Year1998 0.07521 0.17355 0.43 0.66
## Year1999 0.17411 0.18245 0.95 0.34
## Year2000 0.04473 0.16975 0.26 0.79
## Year2001 0.01245 0.15802 0.08 0.94
## Year2002 0.05851 0.15565 0.38 0.71
## Year2003 -0.00647 0.15369 -0.04 0.97
## Year2004 -0.03719 0.15591 -0.24 0.81
## Year2005 -0.04442 0.16343 -0.27 0.79
```

```

## Year2006      -0.00692    0.15127   -0.05    0.96
## Year2007      -0.07959    0.16207   -0.49    0.62
## Year2008      -0.01308    0.15389   -0.09    0.93
## Year2009      -0.09889    0.15446   -0.64    0.52
## Year2010      -0.13171    0.15303   -0.86    0.39
## Year2011      -0.05314    0.15247   -0.35    0.73
## Year2012      -0.05439    0.16766   -0.32    0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
## Multiple R-squared:  0.0464, Adjusted R-squared:  0.0147
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 516 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.210  0.873   0.952   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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.106 1      1.052
## Year      1.106 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
## -1.1758 -0.2249 -0.0122 0.2287 1.0802
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14866 0.14717 7.80 3.1e-14 ***
## FirstAuthorFemale1 -0.02193 0.03196 -0.69 0.49
## Year1997 0.14469 0.17424 0.83 0.41
## Year1998 0.07036 0.17442 0.40 0.69
## Year1999 0.17659 0.18479 0.96 0.34
## Year2000 0.04905 0.17167 0.29 0.78
## Year2001 0.01200 0.15978 0.08 0.94
## Year2002 0.05520 0.15666 0.35 0.72
## Year2003 -0.00460 0.15559 -0.03 0.98
## Year2004 -0.03465 0.15805 -0.22 0.83
## Year2005 -0.04638 0.16435 -0.28 0.78
## Year2006 -0.00603 0.15296 -0.04 0.97
```

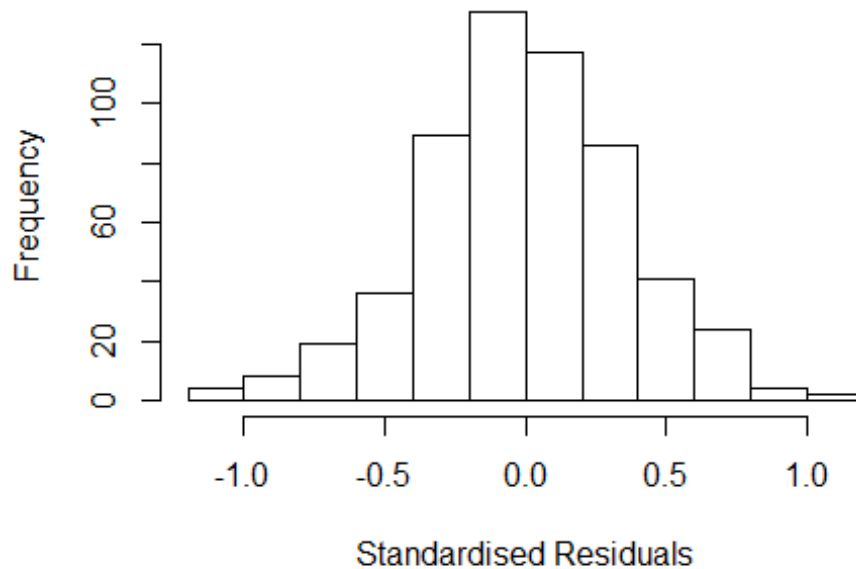


```

## Year2007      -0.08777    0.16322   -0.54    0.59
## Year2008      -0.01669    0.15528   -0.11    0.91
## Year2009      -0.09923    0.15590   -0.64    0.52
## Year2010      -0.13534    0.15441   -0.88    0.38
## Year2011      -0.05956    0.15323   -0.39    0.70
## Year2012      -0.05885    0.16894   -0.35    0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0145
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.212  0.871   0.955   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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.148 1          1.071
## Year              1.148 16          1.004

```

Residuals from last author



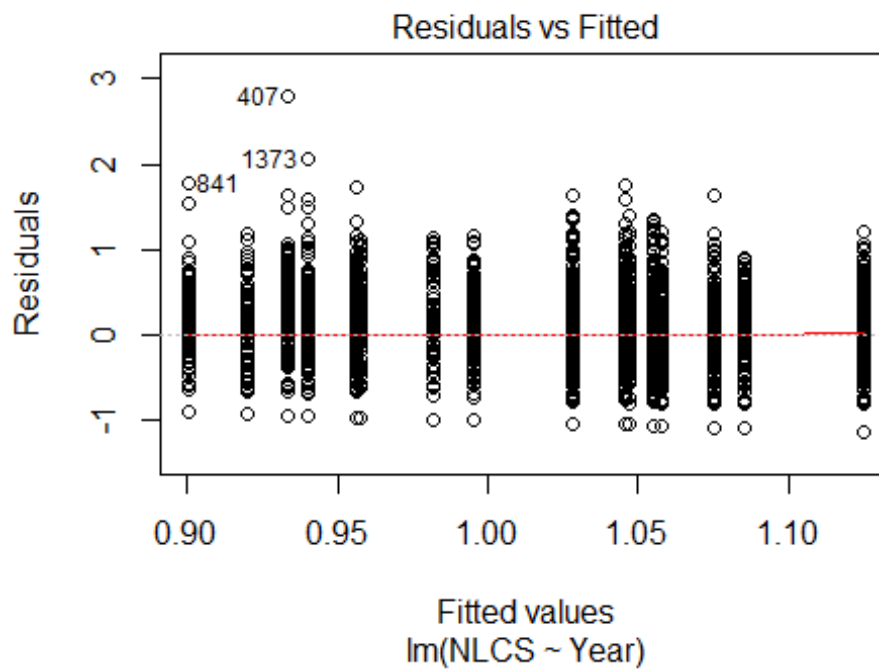
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1955 -0.2275 -0.0133 0.2188 1.0815
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15066 0.14517 7.93 1.3e-14 ***
## LastAuthorFemale1 -0.04311 0.03720 -1.16 0.25
## Year1997 0.14363 0.17349 0.83 0.41
## Year1998 0.07431 0.17419 0.43 0.67
## Year1999 0.17234 0.18335 0.94 0.35
## Year2000 0.04484 0.17006 0.26 0.79
## Year2001 0.00806 0.15804 0.05 0.96
## Year2002 0.05430 0.15597 0.35 0.73
## Year2003 -0.01166 0.15365 -0.08 0.94
## Year2004 -0.04127 0.15610 -0.26 0.79
## Year2005 -0.05058 0.16367 -0.31 0.76
## Year2006 -0.01017 0.15137 -0.07 0.95
```

```

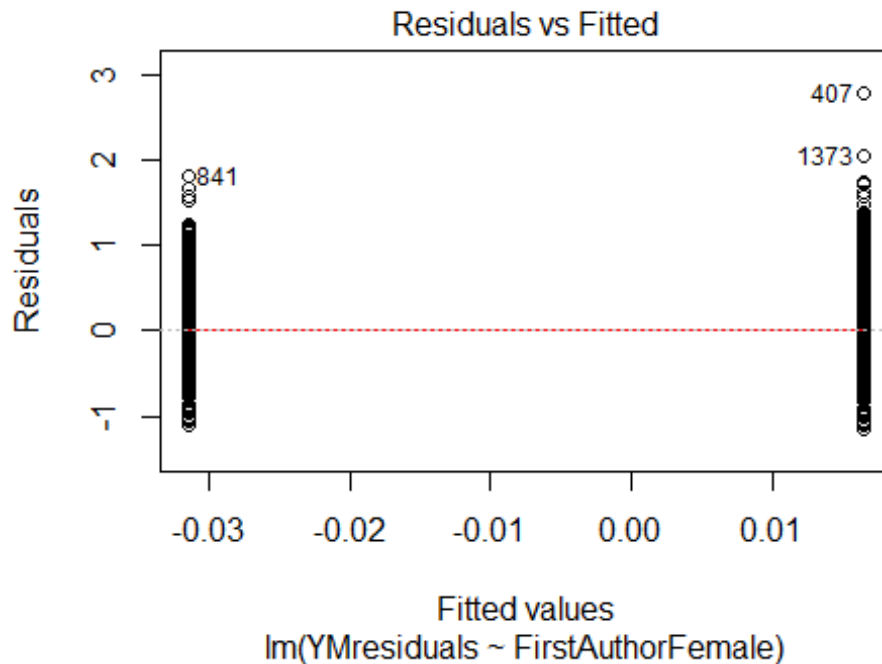
## Year2007          -0.07910      0.16235    -0.49      0.63
## Year2008          -0.01525      0.15414    -0.10      0.92
## Year2009          -0.10194      0.15479    -0.66      0.51
## Year2010          -0.13439      0.15326    -0.88      0.38
## Year2011          -0.05615      0.15267    -0.37      0.71
## Year2012          -0.06215      0.16636    -0.37      0.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.342
## Multiple R-squared:  0.0461, Adjusted R-squared:  0.0162
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 520 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.196  0.873   0.953   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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 561"
## [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
##  146  162  194  201  214  212  243  233  227  255  282  297  329  393  414
## 2011 2012
##  420  422
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   89   99  126  128  124  113  181  172  172  191  207  211  237  297  321
## 2011 2012

```

```
## 314 299
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 93 124 121 119 108 174 164 156 176 196 199 220 277 292
## 2011 2012
## 285 274
## [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
```

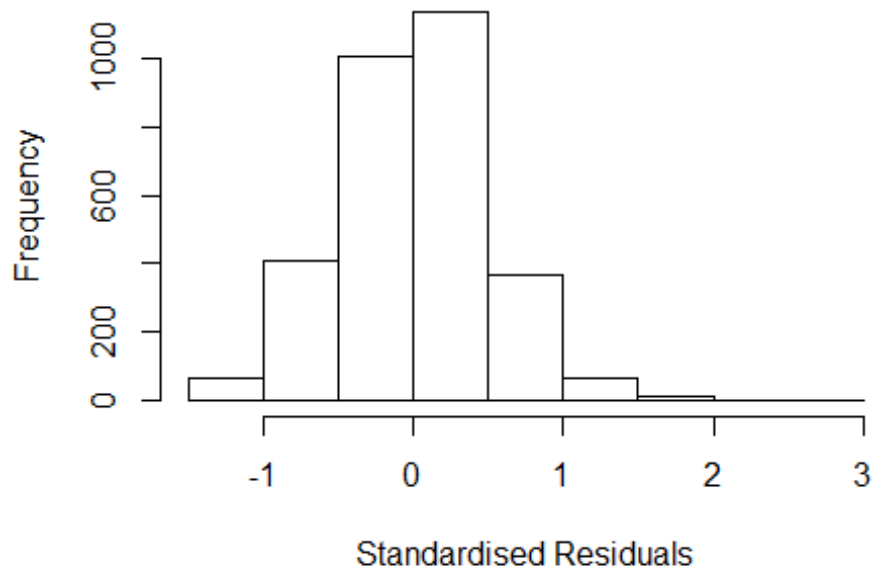


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.37, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 2.52553590965746"
## [1] "Male first author team size 2018 geometric mean: 2.44044604015248"
##
## 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: 2.19774025235148"
## [1] "Male last author team size 2018 geometric mean: 2.63998041454875"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, 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 1.231 1      1.109
## LastAuthorFemale  1.238 1      1.112
## UniqueAuthors    1.184 4      1.021
## Year             1.196 16      1.006
```

Residuals from first and last author and team size



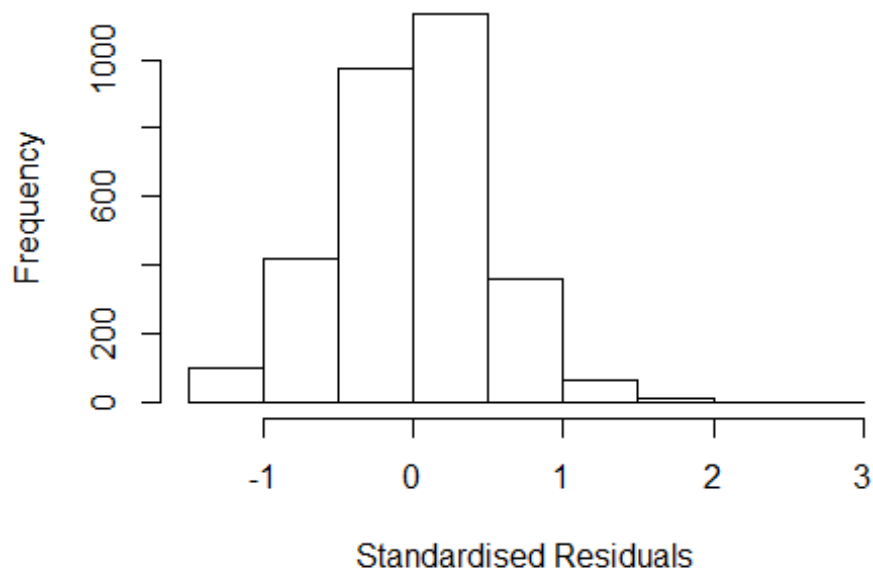
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 407 0032077790 3.725 1998    2100      2    2.882
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2607 -0.3124  0.0144  0.3144  2.8820
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90055    0.05851   15.39 < 2e-16 ***
## FirstAuthorFemale1 -0.03953    0.02124   -1.86  0.063 .
## LastAuthorFemale1 -0.01484    0.02224   -0.67  0.504
## UniqueAuthors2     0.12103    0.02359    5.13 3.1e-07 ***
## UniqueAuthors3     0.16757    0.02510    6.68 2.9e-11 ***
## UniqueAuthors4     0.20403    0.03361    6.07 1.4e-09 ***
## UniqueAuthors5     0.26412    0.03554    7.43 1.4e-13 ***
## Year1997        -0.03816    0.08110   -0.47  0.638
## Year1998        -0.05758    0.08015   -0.72  0.473
## Year1999         0.00788    0.07776    0.10  0.919
```

```

## Year2000      -0.06970    0.07832   -0.89    0.374
## Year2001      0.10297    0.07347    1.40    0.161
## Year2002     -0.05420    0.07369   -0.74    0.462
## Year2003      0.06583    0.06821    0.97    0.335
## Year2004     -0.02425    0.06889   -0.35    0.725
## Year2005     -0.02632    0.06651   -0.40    0.692
## Year2006      0.02746    0.06576    0.42    0.676
## Year2007      0.07120    0.06732    1.06    0.290
## Year2008      0.09604    0.06868    1.40    0.162
## Year2009      0.03065    0.06621    0.46    0.643
## Year2010      0.05948    0.06565    0.91    0.365
## Year2011      0.05479    0.06624    0.83    0.408
## Year2012      0.13380    0.06519    2.05    0.040 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0514, Adjusted R-squared:  0.0445
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 218 is an outlier with |weight| = 0 ( < 3.3e-05);
## 287 weights are ~= 1. The remaining 2771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0003 0.8560 0.9490 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          3.27e-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 1.227 1 1.108
## LastAuthorFemale 1.210 1 1.100
## Year 1.052 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
## 407 0032077790 3.725 1998    2100      2      2.817
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1566 -0.3210  0.0161  0.3180  2.8171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.964860   0.060910   15.84  <2e-16 ***
## FirstAuthorFemale1 -0.029710   0.021599   -1.38   0.1691
## LastAuthorFemale1 -0.036799   0.022515   -1.63   0.1023
## Year1997        -0.040520   0.084277   -0.48   0.6307
## Year1998        -0.056927   0.083482   -0.68   0.4953
## Year1999         0.016739   0.080683    0.21   0.8357
## Year2000        -0.054546   0.082301   -0.66   0.5075
## Year2001         0.134332   0.076859    1.75   0.0806 .
## Year2002        -0.022201   0.075435   -0.29   0.7685
## Year2003         0.091121   0.070469    1.29   0.1961
## Year2004         0.004758   0.071704    0.07   0.9471
## Year2005         0.000284   0.068682    0.00   0.9967
```

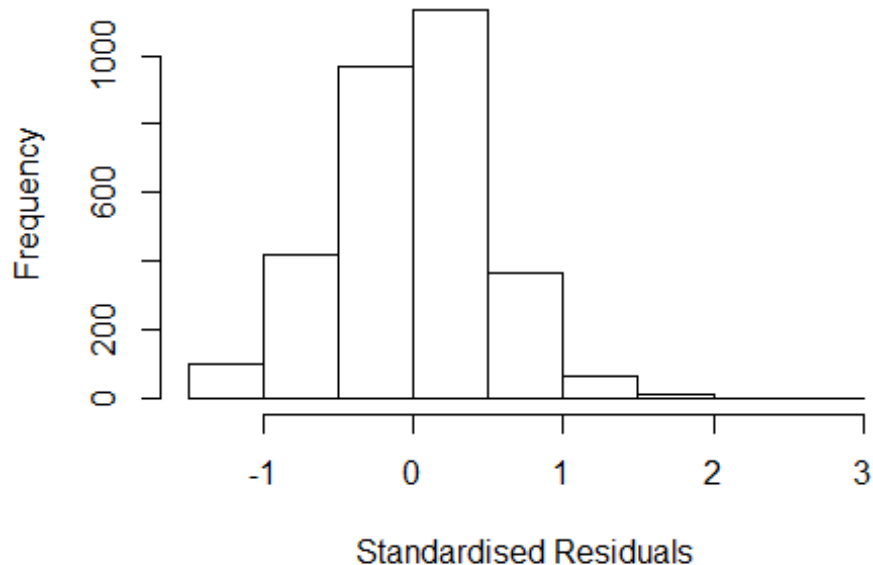


```

## Year2006          0.062756    0.068499    0.92    0.3597
## Year2007          0.123037    0.068860    1.79    0.0741 .
## Year2008          0.137968    0.070435    1.96    0.0502 .
## Year2009          0.077875    0.068139    1.14    0.2532
## Year2010          0.109813    0.067833    1.62    0.1056
## Year2011          0.113777    0.068497    1.66    0.0968 .
## Year2012          0.191786    0.067438    2.84    0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0226, Adjusted R-squared:  0.0168
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 218 is an outlier with |weight| = 0 ( < 3.3e-05);
## 278 weights are ~ = 1. The remaining 2780 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0166 0.8570 0.9470 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      3.27e-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 1.034 1          1.017
## Year              1.034 16          1.001

```

Residuals from first author



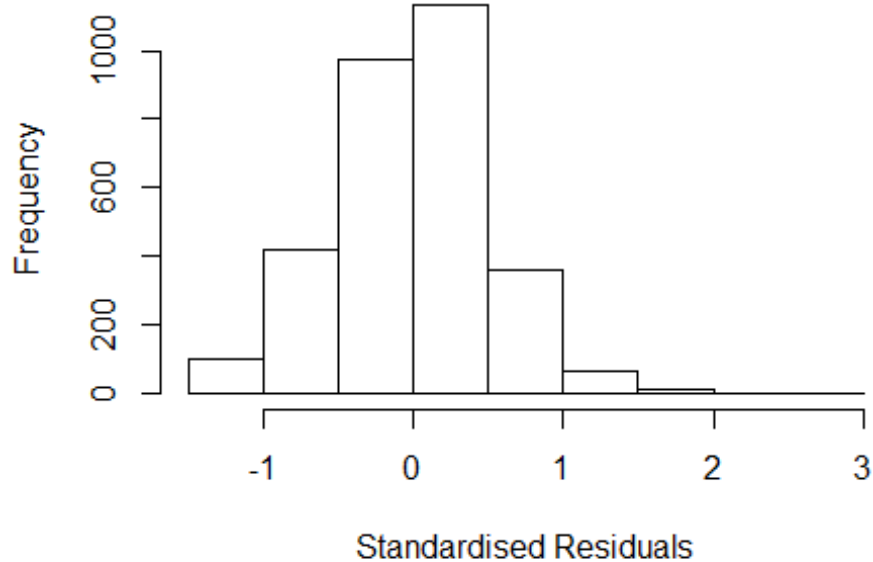
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 407 0032077790 3.725 1998    2100      2    2.817
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1518 -0.3242  0.0164  0.3164  2.8210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95823    0.06090   15.74  <2e-16 ***
## FirstAuthorFemale1 -0.04596    0.01993   -2.31   0.0212 *
## Year1997       -0.03960    0.08438   -0.47   0.6389
## Year1998       -0.05422    0.08360   -0.65   0.5166
## Year1999        0.01955    0.08072    0.24   0.8086
## Year2000       -0.05456    0.08234   -0.66   0.5076
## Year2001        0.13711    0.07689    1.78   0.0746 .
## Year2002       -0.02182    0.07552   -0.29   0.7727
## Year2003        0.09535    0.07052    1.35   0.1764
## Year2004        0.00731    0.07178    0.10   0.9189
## Year2005        0.00200    0.06881    0.03   0.9768
## Year2006        0.06477    0.06867    0.94   0.3457
```

```

## Year2007          0.12337      0.06904      1.79      0.0741 .
## Year2008          0.13912      0.07055      1.97      0.0487 *
## Year2009          0.07945      0.06825      1.16      0.2445
## Year2010          0.11035      0.06797      1.62      0.1046
## Year2011          0.11367      0.06866      1.66      0.0979 .
## Year2012          0.19361      0.06756      2.87      0.0042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0217, Adjusted R-squared:  0.0163
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 218 is an outlier with |weight| = 0 ( < 3.3e-05);
## 278 weights are ~= 1. The remaining 2780 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0154 0.8570 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      3.27e-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          1.010
## Year              1.019 16          1.001

```

Residuals from last author



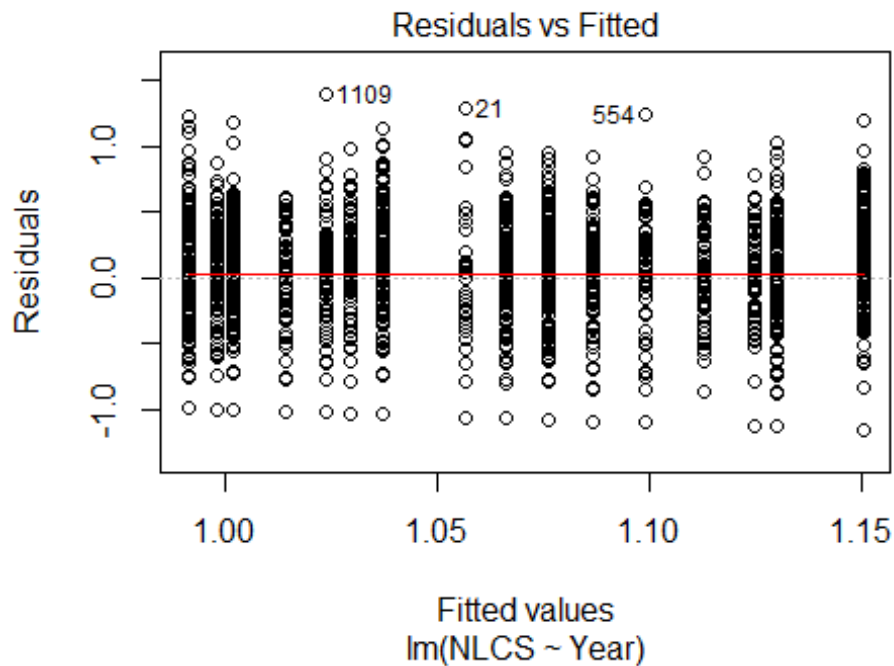
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 407 0032077790 3.725 1998    2100      2    2.817
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1507 -0.3253  0.0177  0.3178  2.8220
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96054    0.06111   15.72  <2e-16 ***
## LastAuthorFemale1 -0.05105    0.02077   -2.46   0.014 *
## Year1997      -0.04031    0.08438   -0.48   0.633
## Year1998      -0.05753    0.08391   -0.69   0.493
## Year1999       0.01624    0.08104    0.20   0.841
## Year2000      -0.05360    0.08257   -0.65   0.516
## Year2001       0.13171    0.07702    1.71   0.087 .
## Year2002      -0.02211    0.07566   -0.29   0.770
## Year2003       0.08861    0.07060    1.26   0.210
## Year2004       0.00380    0.07201    0.05   0.958
## Year2005      -0.00134    0.06890   -0.02   0.984
## Year2006       0.05927    0.06866    0.86   0.388
```

```

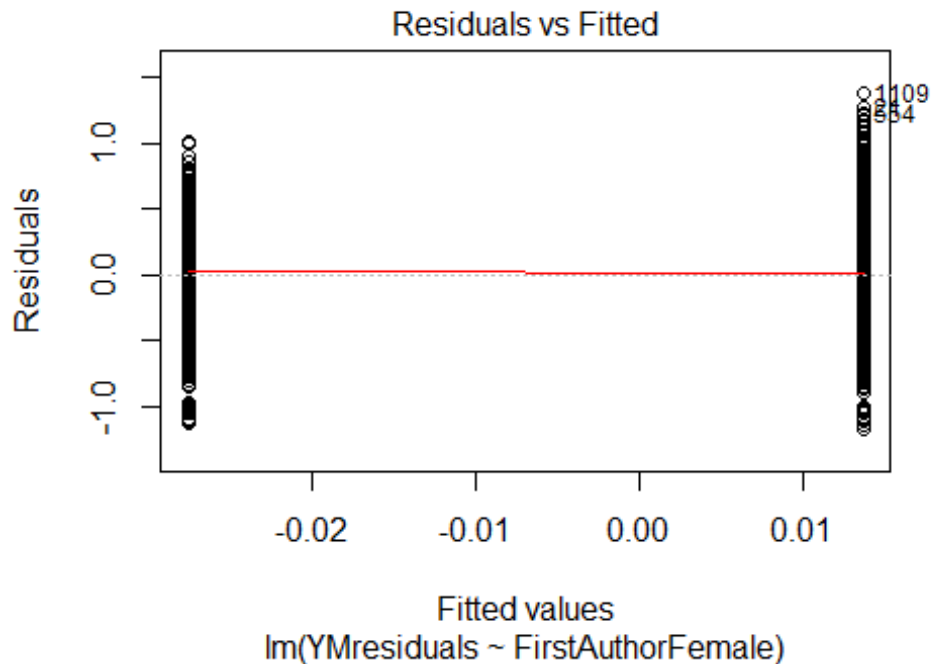
## Year2007      0.12084      0.06902      1.75      0.080 .
## Year2008      0.13405      0.07062      1.90      0.058 .
## Year2009      0.07470      0.06831      1.09      0.274
## Year2010      0.10822      0.06800      1.59      0.112
## Year2011      0.11272      0.06871      1.64      0.101
## Year2012      0.19013      0.06764      2.81      0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.0164
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 218 is an outlier with |weight| = 0 ( < 3.3e-05);
## 269 weights are ~= 1. The remaining 2789 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0155 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      3.27e-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: 3059"
## [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
## 101 99 118 114 131 144 161 140 132 135 176 173 200 259 265
## 2011 2012
## 245 263
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 39 63 76 60 67 65 111 90 95 86 129 128 155 197 203

```

```
## 2011 2012
## 192 200
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 63 71 57 62 60 106 83 85 78 123 117 141 184 180
## 2011 2012
## 177 178
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```

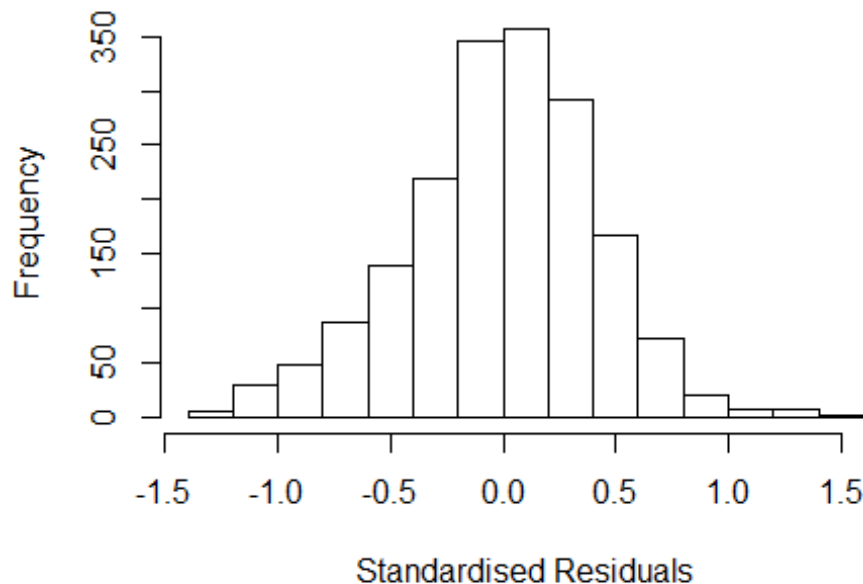


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.83, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.19801509081169"
## [1] "Male first author team size 2018 geometric mean: 3.57148114345696"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.0068126907536"
## [1] "Male last author team size 2018 geometric mean: 3.57441504019536"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3700, 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 1.092 1      1.045
## LastAuthorFemale  1.094 1      1.046
## UniqueAuthors     1.310 4      1.034
## Year              1.356 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
## -1.3440 -0.2719 0.0178 0.2720 1.4055
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96218 0.07283 13.21 < 2e-16 ***
## FirstAuthorFemale1 -0.03751 0.02247 -1.67 0.09514 .
## LastAuthorFemale1 -0.03690 0.02517 -1.47 0.14274
## UniqueAuthors2 0.11743 0.03102 3.79 0.00016 ***
## UniqueAuthors3 0.19951 0.03157 6.32 3.3e-10 ***
## UniqueAuthors4 0.29538 0.03709 7.96 3.0e-15 ***
## UniqueAuthors5 0.26953 0.03824 7.05 2.6e-12 ***
## Year1997 -0.00498 0.08618 -0.06 0.95397
## Year1998 -0.08311 0.08736 -0.95 0.34159
## Year1999 0.05105 0.08758 0.58 0.56007
```

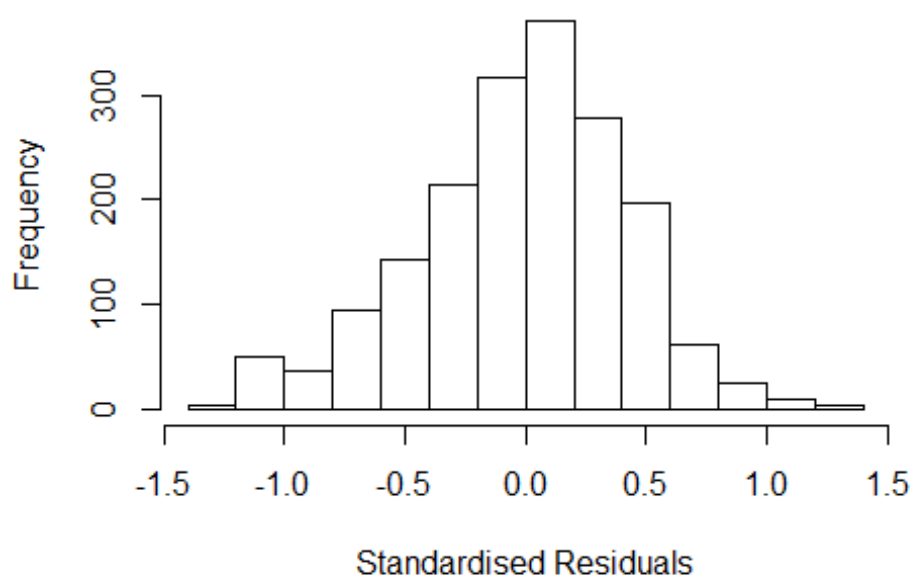


```

## Year2000      0.04289    0.09239    0.46  0.64254
## Year2001      0.08643    0.08706    0.99  0.32094
## Year2002     -0.07673    0.08980   -0.85  0.39297
## Year2003     -0.06911    0.08297   -0.83  0.40501
## Year2004     -0.05074    0.08328   -0.61  0.54238
## Year2005      0.01633    0.08523    0.19  0.84808
## Year2006      0.03652    0.08239    0.44  0.65763
## Year2007     -0.09205    0.08124   -1.13  0.25733
## Year2008     -0.01975    0.08114   -0.24  0.80773
## Year2009     -0.07862    0.07988   -0.98  0.32510
## Year2010     -0.00086    0.07934   -0.01  0.99136
## Year2011     -0.01915    0.08225   -0.23  0.81592
## Year2012      0.04729    0.08028    0.59  0.55587
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0747, Adjusted R-squared:  0.0633
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1652 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.197  0.865   0.950   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      5.55e-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 1.089 1      1.043
## LastAuthorFemale  1.082 1      1.040
## Year              1.080 16      1.002

```

Residuals from first and last author



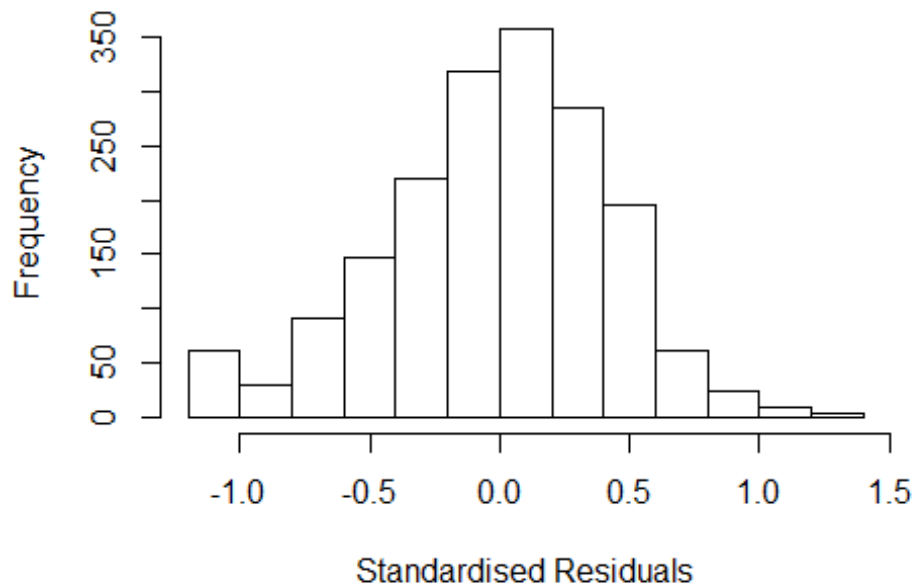
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2003 -0.2840  0.0173  0.2813  1.3897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.046152   0.074968   13.95  <2e-16 ***
## FirstAuthorFemale1 -0.024226   0.022891   -1.06   0.290
## LastAuthorFemale1 -0.054271   0.025359   -2.14   0.032 *
## Year1997         0.005043   0.089321    0.06   0.955
## Year1998        -0.066754   0.090421   -0.74   0.460
## Year1999         0.068788   0.090585    0.76   0.448
## Year2000         0.107147   0.097423    1.10   0.272
## Year2001         0.108629   0.091015    1.19   0.233
## Year2002        -0.022616   0.091937   -0.25   0.806
## Year2003        -0.019844   0.086161   -0.23   0.818
## Year2004         0.006793   0.087896    0.08   0.938
## Year2005         0.062454   0.087724    0.71   0.477
```

```

## Year2006          0.122465    0.084845    1.44    0.149
## Year2007         -0.010210    0.083742   -0.12    0.903
## Year2008          0.055417    0.082600    0.67    0.502
## Year2009          0.000239    0.081880    0.00    0.998
## Year2010          0.083780    0.081558    1.03    0.304
## Year2011          0.048573    0.084947    0.57    0.568
## Year2012          0.154187    0.082088    1.88    0.061 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.0125
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1642 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  0.865   0.946   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      5.55e-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 1.045 1      1.022
## Year              1.045 16      1.001

```

Residuals from first author



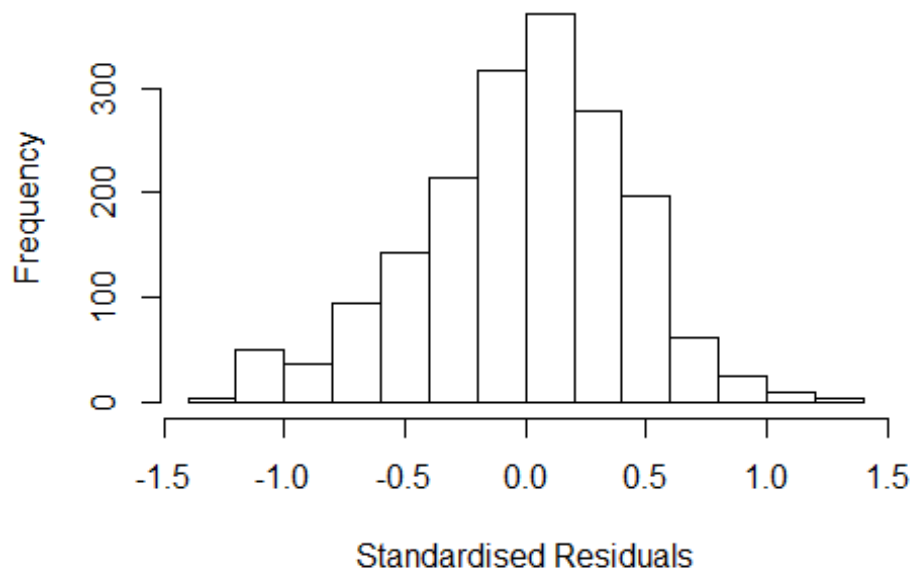
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1894 -0.2817 0.0217 0.2836 1.3963
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04e+00 7.45e-02 13.95 <2e-16 ***
## FirstAuthorFemale1 -3.64e-02 2.26e-02 -1.61 0.108
## Year1997 8.63e-04 8.91e-02 0.01 0.992
## Year1998 -6.30e-02 9.02e-02 -0.70 0.485
## Year1999 7.21e-02 9.02e-02 0.80 0.424
## Year2000 1.05e-01 9.74e-02 1.08 0.280
## Year2001 1.03e-01 9.09e-02 1.13 0.258
## Year2002 -2.42e-02 9.19e-02 -0.26 0.792
## Year2003 -1.95e-02 8.58e-02 -0.23 0.821
## Year2004 7.03e-03 8.78e-02 0.08 0.936
## Year2005 6.10e-02 8.75e-02 0.70 0.486
## Year2006 1.23e-01 8.47e-02 1.45 0.146
```

```

## Year2007          -1.59e-02   8.36e-02   -0.19   0.849
## Year2008          5.29e-02   8.22e-02    0.64   0.520
## Year2009          8.95e-05   8.15e-02    0.00   0.999
## Year2010          8.17e-02   8.14e-02    1.00   0.316
## Year2011          3.98e-02   8.47e-02    0.47   0.638
## Year2012          1.50e-01   8.18e-02    1.84   0.066 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0102
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1643 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.864  0.946  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      5.55e-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 1.037 1          1.018
## Year            1.037 16          1.001

```

Residuals from last author



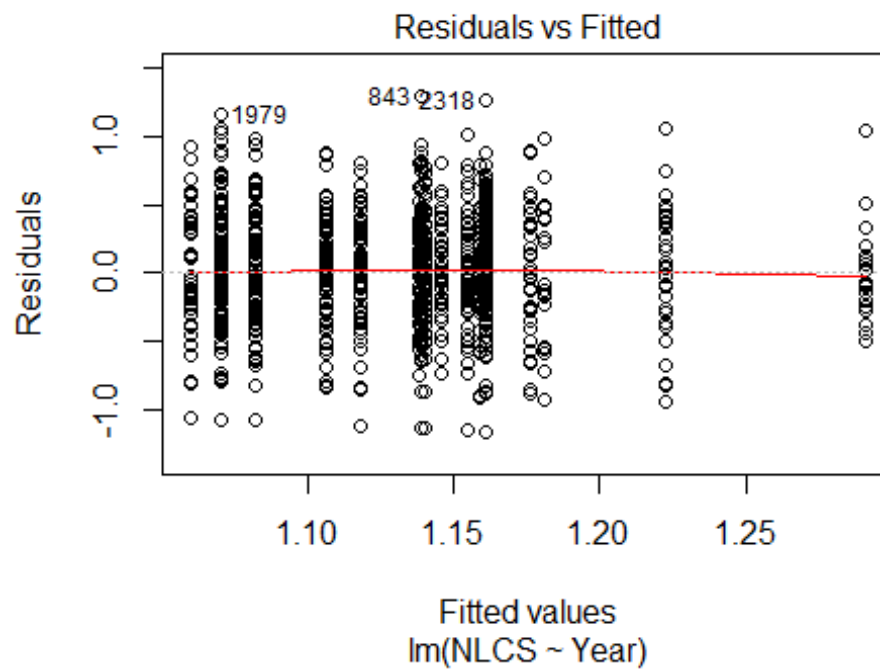
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1927 -0.2815 0.0203 0.2818 1.3941
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04342 0.07487 13.94 <2e-16 ***
## LastAuthorFemale1 -0.06073 0.02495 -2.43 0.015 *
## Year1997 0.00521 0.08933 0.06 0.953
## Year1998 -0.06798 0.09036 -0.75 0.452
## Year1999 0.06940 0.09049 0.77 0.443
## Year2000 0.10526 0.09708 1.08 0.278
## Year2001 0.10528 0.09093 1.16 0.247
## Year2002 -0.02406 0.09198 -0.26 0.794
## Year2003 -0.02149 0.08598 -0.25 0.803
## Year2004 0.00355 0.08778 0.04 0.968
## Year2005 0.05918 0.08759 0.68 0.499
## Year2006 0.11791 0.08458 1.39 0.163
```

```

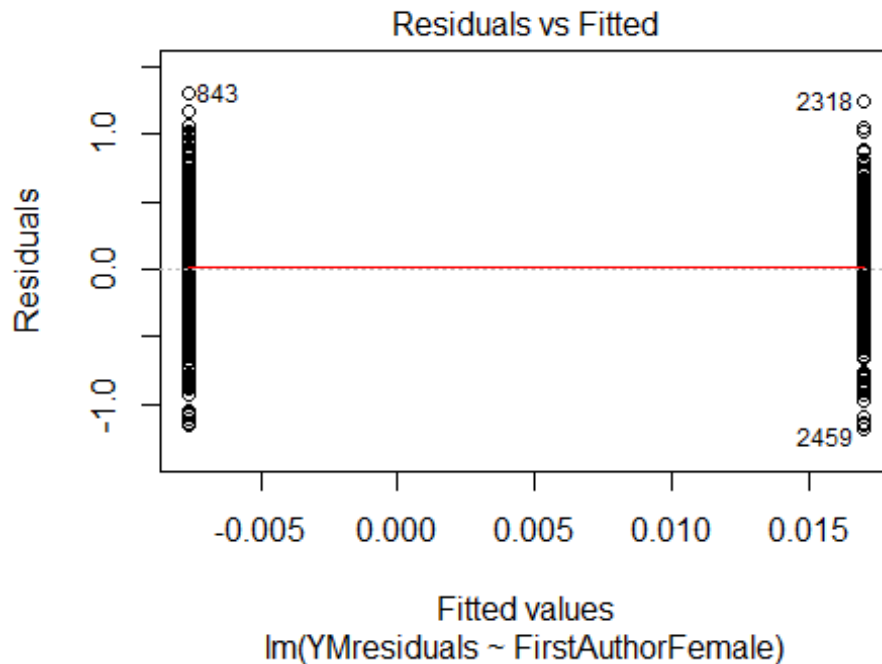
## Year2007          -0.01336      0.08357    -0.16      0.873
## Year2008           0.04988      0.08243      0.61      0.545
## Year2009          -0.00658      0.08155    -0.08      0.936
## Year2010           0.07922      0.08139      0.97      0.330
## Year2011           0.04416      0.08490      0.52      0.603
## Year2012           0.14932      0.08180      1.83      0.068 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0217, Adjusted R-squared:  0.0124
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1643 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.228  0.865  0.945  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      5.55e-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: 1803"
## [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
##   47   80   59   81   99   96   88   76   93   85  118  102  123  146  180
## 2011 2012
##  233  243
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   40   19   36   34   41   49   41   47   42   71   63   68   82  114
## 2011 2012

```

```
## 132 142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 36 17 35 32 37 46 36 38 36 59 52 63 71 99
## 2011 2012
## 106 123
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```

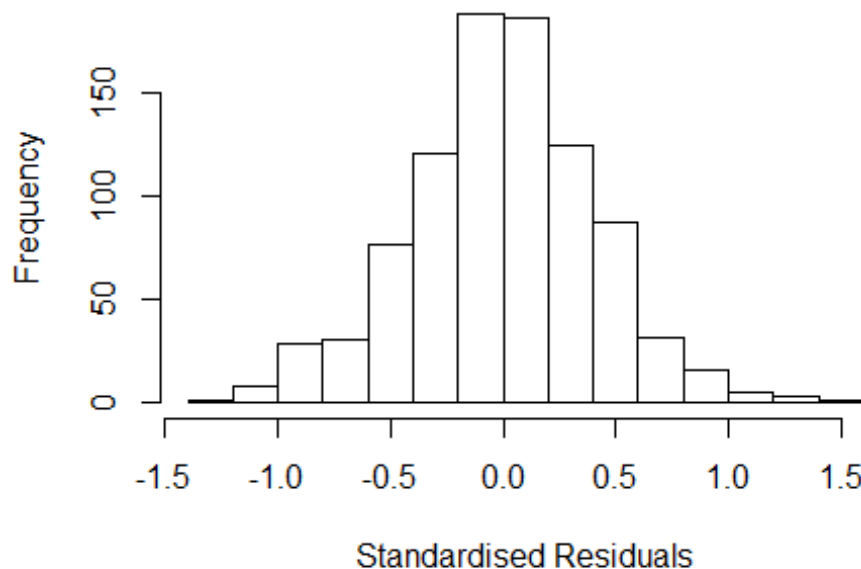


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.04, df = 1, p-value = 0.8
```

```
## [1] "Female first author team size 2018 geometric mean: 4.50993363802035"
## [1] "Male first author team size 2018 geometric mean: 3.73232312742578"
##
## 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: 4.05380410184181"
## [1] "Male last author team size 2018 geometric mean: 3.99386232294942"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, 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 1.160 1      1.077
## LastAuthorFemale  1.098 1      1.048
## UniqueAuthors    1.560 4      1.057
## Year              1.888 16      1.020
```

Residuals from first and last author and team size



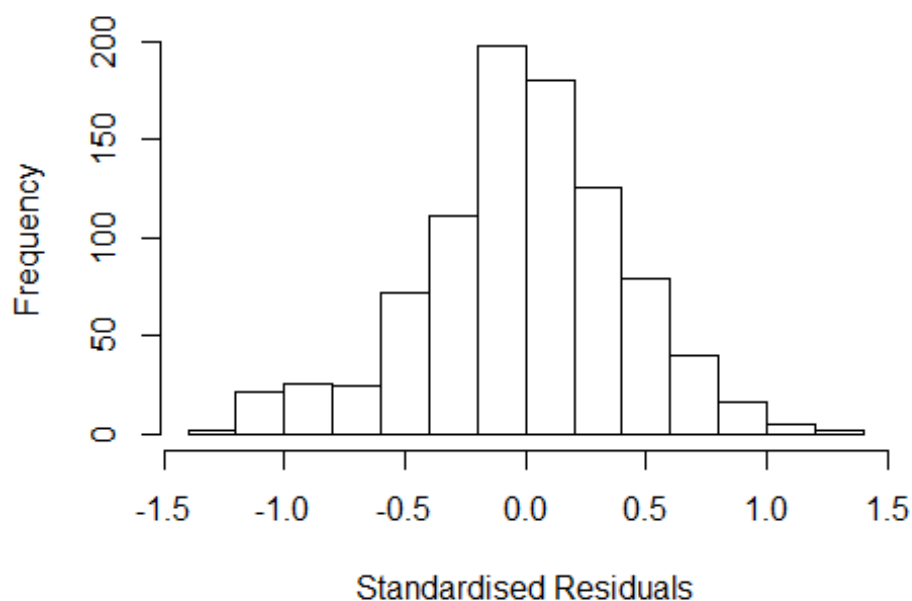
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24128 -0.24567  0.00288  0.26706  1.40903
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0267    0.0989   10.38 < 2e-16 ***
## FirstAuthorFemale1  0.0251    0.0311    0.81  0.41878
## LastAuthorFemale1  0.0438    0.0328    1.34  0.18156
## UniqueAuthors2    0.2192    0.0611    3.59  0.00035 ***
## UniqueAuthors3    0.2905    0.0631    4.60  4.8e-06 ***
## UniqueAuthors4    0.2651    0.0630    4.21  2.8e-05 ***
## UniqueAuthors5    0.3770    0.0632    5.96  3.6e-09 ***
## Year1997         -0.0654    0.1231   -0.53  0.59540
## Year1998         -0.1097    0.1785   -0.61  0.53888
## Year1999         -0.1203    0.1340   -0.90  0.36941
```

```

## Year2000          -0.0344      0.1267   -0.27   0.78629
## Year2001          -0.2251      0.1201   -1.87   0.06124 .
## Year2002          -0.0915      0.1067   -0.86   0.39164
## Year2003          -0.1636      0.1139   -1.44   0.15121
## Year2004          -0.2250      0.1111   -2.03   0.04315 *
## Year2005          -0.1130      0.1137   -0.99   0.32054
## Year2006          -0.1497      0.1059   -1.41   0.15788
## Year2007          -0.1302      0.1092   -1.19   0.23347
## Year2008          -0.1349      0.1076   -1.25   0.21015
## Year2009          -0.1720      0.1039   -1.66   0.09819 .
## Year2010          -0.2062      0.1039   -1.98   0.04755 *
## Year2011          -0.2395      0.1024   -2.34   0.01960 *
## Year2012          -0.1461      0.1029   -1.42   0.15599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.088, Adjusted R-squared:  0.0652
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 827 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.870  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          1.11e-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 1.184 1          1.088
## LastAuthorFemale  1.065 1          1.032
## Year              1.260 16          1.007

```

Residuals from first and last author



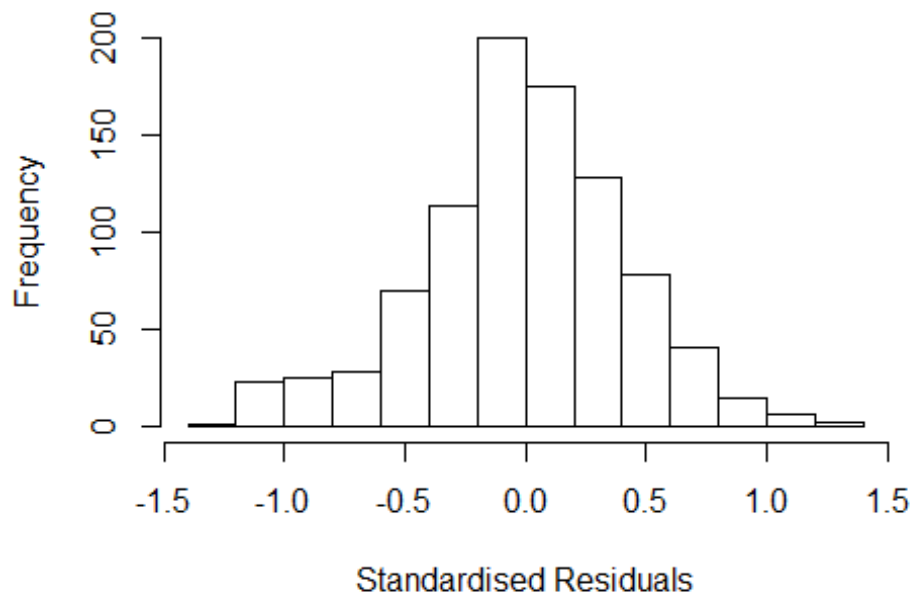
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23972 -0.25271 -0.00526  0.26633  1.36441
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2297    0.0663   18.55  <2e-16 ***
## FirstAuthorFemale1  0.0377    0.0320    1.18   0.240
## LastAuthorFemale1  0.0330    0.0331    1.00   0.319
## Year1997          -0.0665    0.1018   -0.65   0.514
## Year1998          -0.0820    0.1718   -0.48   0.633
## Year1999          -0.1211    0.1192   -1.02   0.310
## Year2000          -0.0107    0.1116   -0.10   0.924
## Year2001          -0.2262    0.1060   -2.13   0.033 *
## Year2002          -0.0615    0.0862   -0.71   0.476
## Year2003          -0.1016    0.0927   -1.10   0.273
## Year2004          -0.1641    0.0943   -1.74   0.082 .
## Year2005          -0.0876    0.0908   -0.96   0.335
```

```

## Year2006          -0.0987      0.0850   -1.16    0.246
## Year2007          -0.0876      0.0858   -1.02    0.308
## Year2008          -0.0856      0.0876   -0.98    0.329
## Year2009          -0.1120      0.0816   -1.37    0.171
## Year2010          -0.1465      0.0809   -1.81    0.070 .
## Year2011          -0.1818      0.0812   -2.24    0.025 *
## Year2012          -0.0606      0.0787   -0.77    0.441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0179, Adjusted R-squared:  -0.00211
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 820 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.206  0.858  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      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 1.174 1      1.083
## Year              1.174 16      1.005

```

Residuals from first author



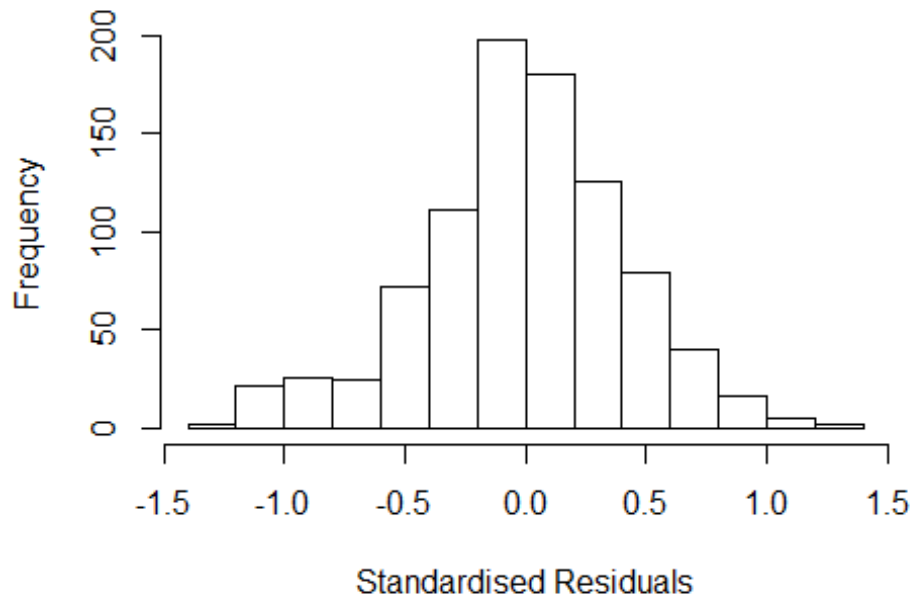
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21791 -0.25421 -0.00822 0.27161 1.35868
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22952 0.06638 18.52 <2e-16 ***
## FirstAuthorFemale1 0.04199 0.03210 1.31 0.191
## Year1997 -0.06284 0.10168 -0.62 0.537
## Year1998 -0.08165 0.17089 -0.48 0.633
## Year1999 -0.11634 0.11945 -0.97 0.330
## Year2000 -0.00331 0.11167 -0.03 0.976
## Year2001 -0.22153 0.10597 -2.09 0.037 *
## Year2002 -0.05039 0.08550 -0.59 0.556
## Year2003 -0.10079 0.09296 -1.08 0.279
## Year2004 -0.15820 0.09451 -1.67 0.095 .
## Year2005 -0.08195 0.09039 -0.91 0.365
## Year2006 -0.09090 0.08504 -1.07 0.285
```

```

## Year2007          -0.07772    0.08518   -0.91    0.362
## Year2008          -0.07787    0.08758   -0.89    0.374
## Year2009          -0.10588    0.08134   -1.30    0.193
## Year2010          -0.14109    0.08075   -1.75    0.081 .
## Year2011          -0.17508    0.08080   -2.17    0.031 *
## Year2012          -0.05360    0.07835   -0.68    0.494
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0168, Adjusted R-squared:  -0.00206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 824 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.859  0.952  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      1.11e-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 1.063 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
## -1.21696 -0.25831 -0.00283  0.26679  1.34802
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23007    0.06514   18.88  <2e-16 ***
## LastAuthorFemale1 0.03890    0.03321    1.17   0.242
## Year1997      -0.05884    0.10158   -0.58   0.563
## Year1998      -0.07052    0.16923   -0.42   0.677
## Year1999      -0.10613    0.11712   -0.91   0.365
## Year2000      -0.00919    0.11005   -0.08   0.933
## Year2001      -0.22098    0.10521   -2.10   0.036 *
## Year2002      -0.05488    0.08469   -0.65   0.517
## Year2003      -0.08451    0.08910   -0.95   0.343
## Year2004      -0.14809    0.09149   -1.62   0.106
## Year2005      -0.07373    0.08837   -0.83   0.404
## Year2006      -0.09137    0.08310   -1.10   0.272
```

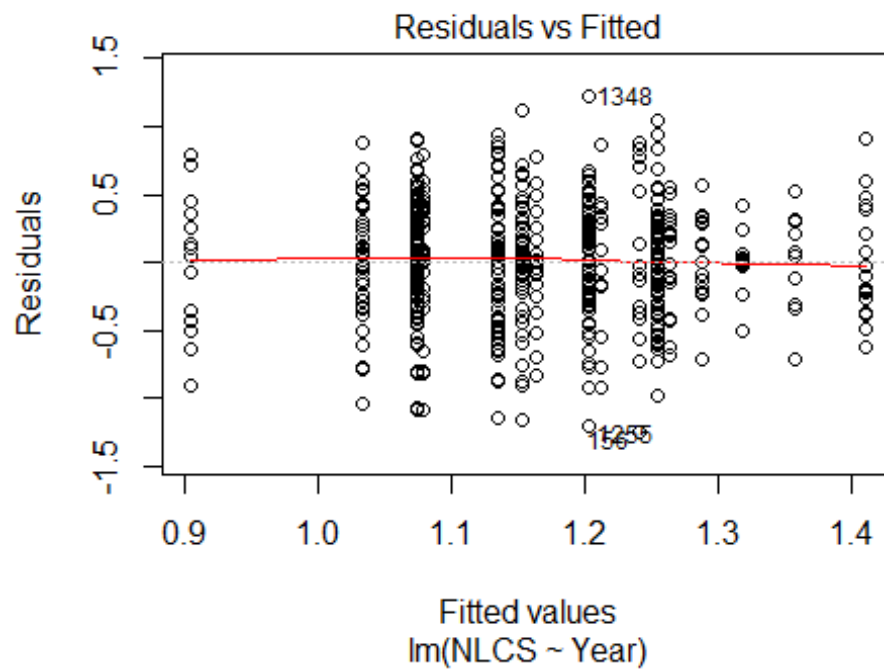


```

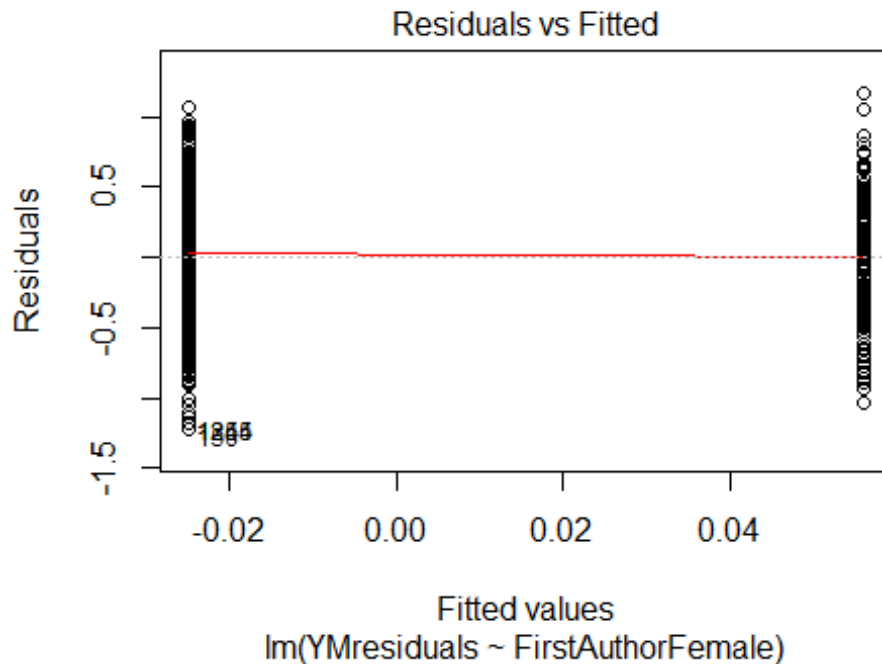
## Year2007          -0.07899      0.08373   -0.94      0.346
## Year2008          -0.07590      0.08500   -0.89      0.372
## Year2009          -0.10210      0.07949   -1.28      0.199
## Year2010          -0.13635      0.07920   -1.72      0.085 .
## Year2011          -0.16848      0.07858   -2.14      0.032 *
## Year2012          -0.05201      0.07701   -0.68      0.500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.016, Adjusted R-squared:  -0.00284
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 813 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.216  0.856  0.951  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      1.11e-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: 904"
## [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
##   34   37   34   34   37   57   42   43   26   42   55   65   89  127  126
## 2011 2012
##  149  145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   14   10   15   14   26   18   19   10   19   34   39   51   73   65
## 2011 2012

```

```
##      82    71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   13    9   14   14   24   18   18    8   17   30   28   46   63   56
## 2011 2012
##    71    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 = 0.06
```

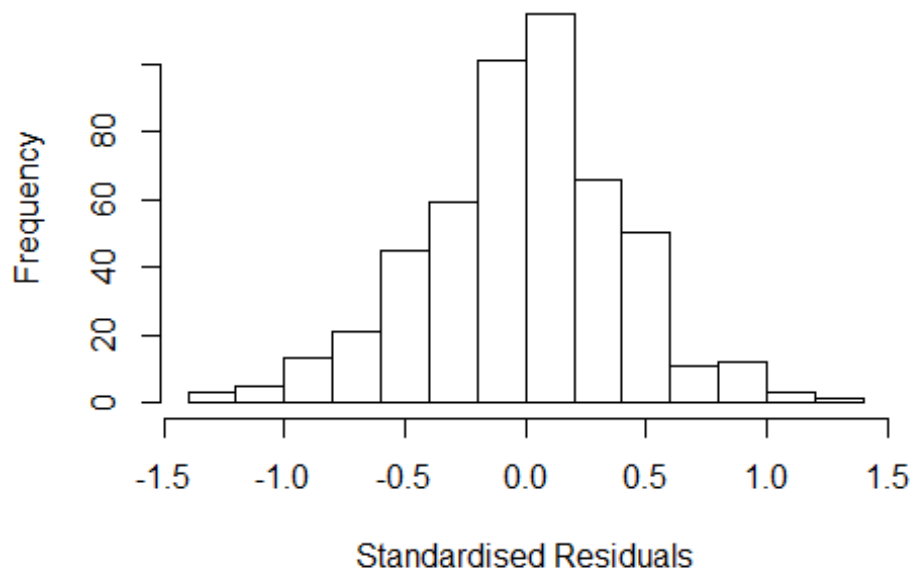


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.3, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.80217920320114"
## [1] "Male first author team size 2018 geometric mean: 4.02564978917357"
##
## 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] "Female last author team size 2018 geometric mean: 3.9846125237134"
## [1] "Male last author team size 2018 geometric mean: 3.92472182754269"
##
## 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 1.189 1      1.091
## LastAuthorFemale  1.190 1      1.091
## UniqueAuthors     1.818 4      1.078
## Year              2.372 16     1.027
```

Residuals from first and last author and team size



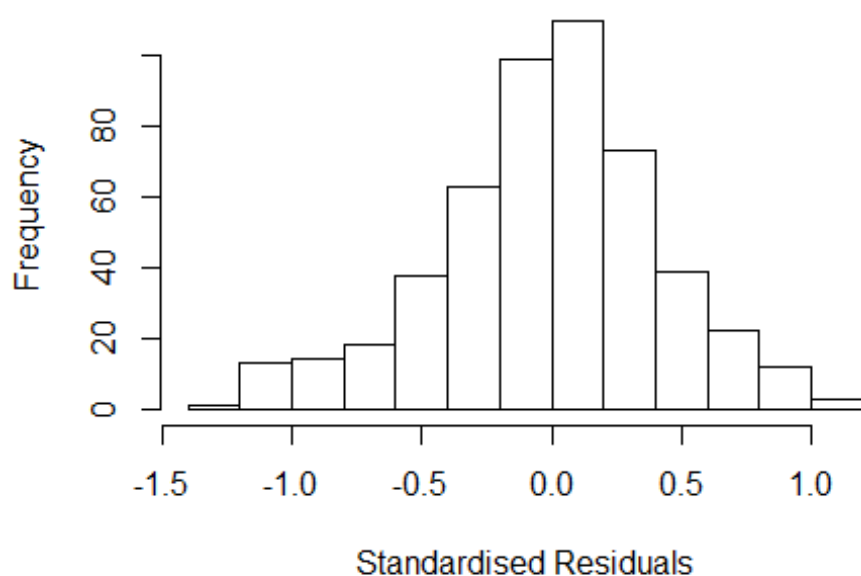
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2629 -0.2380 0.0148 0.2335 1.2717
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0325 0.1225 8.43 4.0e-16 ***
## FirstAuthorFemale1 0.0783 0.0389 2.01 0.04485 *
## LastAuthorFemale1 0.0620 0.0465 1.33 0.18365
## UniqueAuthors2 0.2750 0.0760 3.62 0.00032 ***
## UniqueAuthors3 0.3082 0.0764 4.04 6.3e-05 ***
## UniqueAuthors4 0.4113 0.0778 5.28 1.9e-07 ***
## UniqueAuthors5 0.3895 0.0784 4.97 9.5e-07 ***
## Year1997 -0.0642 0.1694 -0.38 0.70466
## Year1998 -0.0397 0.1765 -0.23 0.82197
## Year1999 -0.1037 0.2146 -0.48 0.62909
```

```

## Year2000          -0.4242      0.1734    -2.45   0.01476 *
## Year2001          -0.2503      0.1381    -1.81   0.07049 .
## Year2002          -0.0543      0.1332    -0.41   0.68370
## Year2003          -0.2375      0.1345    -1.77   0.07802 .
## Year2004          -0.1009      0.1315    -0.77   0.44318
## Year2005          -0.0748      0.1504    -0.50   0.61928
## Year2006          -0.1906      0.1451    -1.31   0.18968
## Year2007          -0.3066      0.1263    -2.43   0.01559 *
## Year2008          -0.1784      0.1299    -1.37   0.17017
## Year2009          -0.2185      0.1199    -1.82   0.06908 .
## Year2010          -0.2319      0.1308    -1.77   0.07687 .
## Year2011          -0.1822      0.1184    -1.54   0.12430
## Year2012          -0.1590      0.1206    -1.32   0.18799
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.135, Adjusted R-squared:  0.0959
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 464 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.859  0.955  0.895  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      1.98e-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 1.212 1      1.101
## LastAuthorFemale  1.168 1      1.081
## Year              1.410 16      1.011

```

Residuals from first and last author



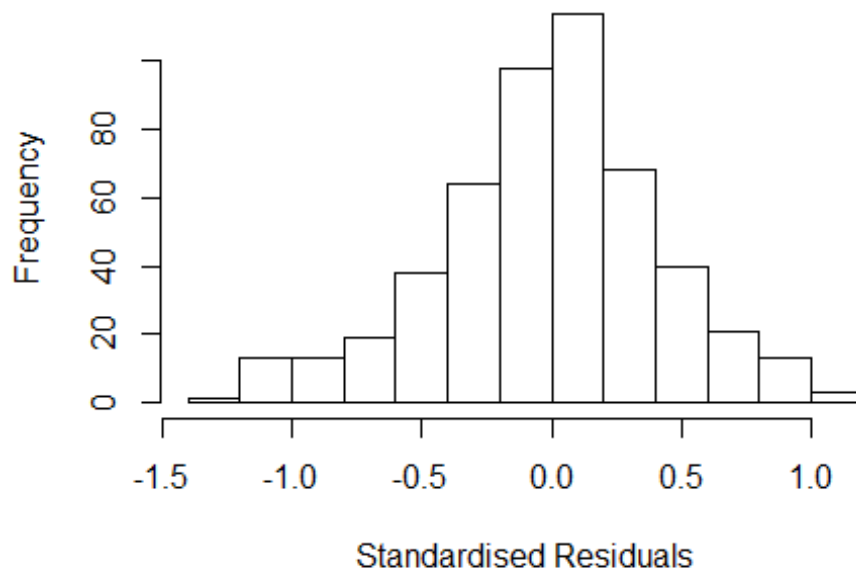
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2233 -0.2611  0.0202  0.2384  1.1122
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3125     0.0851   15.42  <2e-16 ***
## FirstAuthorFemale1  0.0876     0.0400    2.19   0.029 *
## LastAuthorFemale1  0.0378     0.0478    0.79   0.429
## Year1997          -0.0786     0.1640   -0.48   0.632
## Year1998          -0.0146     0.1637   -0.09   0.929
## Year1999          -0.1818     0.2195   -0.83   0.408
## Year2000          -0.4210     0.1741   -2.42   0.016 *
## Year2001          -0.2763     0.1278   -2.16   0.031 *
## Year2002          -0.0665     0.1192   -0.56   0.577
## Year2003          -0.2104     0.1290   -1.63   0.104
## Year2004          -0.0566     0.1154   -0.49   0.624
## Year2005          -0.0563     0.1275   -0.44   0.659
```

```

## Year2006          -0.2127      0.1367    -1.56     0.120
## Year2007          -0.2660      0.1124    -2.37     0.018 *
## Year2008          -0.1917      0.1116    -1.72     0.086 .
## Year2009          -0.2188      0.1037    -2.11     0.035 *
## Year2010          -0.1860      0.1186    -1.57     0.117
## Year2011          -0.1222      0.1001    -1.22     0.223
## Year2012          -0.0892      0.1020    -0.88     0.382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0497, Adjusted R-squared:  0.0145
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 464 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.318  0.857   0.956   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      1.98e-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 1.213 1      1.101
## Year              1.213 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
## -1.2325 -0.2631 0.0129 0.2376 1.1040
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3124 0.0851 15.43 <2e-16 ***
## FirstAuthorFemale1 0.0928 0.0408 2.27 0.023 *
## Year1997 -0.0761 0.1643 -0.46 0.644
## Year1998 -0.0164 0.1644 -0.10 0.920
## Year1999 -0.1740 0.2197 -0.79 0.429
## Year2000 -0.4211 0.1754 -2.40 0.017 *
## Year2001 -0.2734 0.1273 -2.15 0.032 *
## Year2002 -0.0491 0.1170 -0.42 0.675
## Year2003 -0.2101 0.1289 -1.63 0.104
## Year2004 -0.0575 0.1155 -0.50 0.619
## Year2005 -0.0502 0.1262 -0.40 0.691
## Year2006 -0.2020 0.1358 -1.49 0.137
```

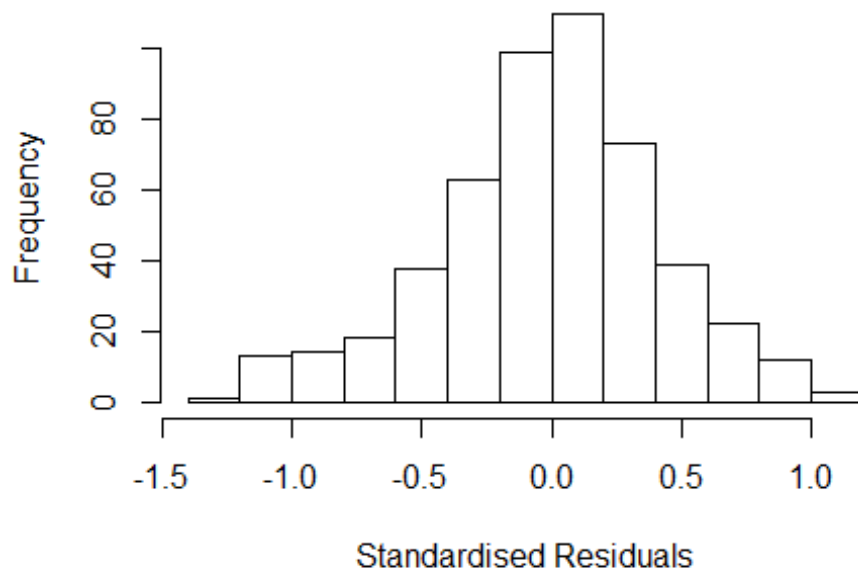


```

## Year2007          -0.2562      0.1126    -2.27      0.023 *
## Year2008          -0.1863      0.1118    -1.67      0.096 .
## Year2009          -0.2090      0.1025    -2.04      0.042 *
## Year2010          -0.1771      0.1176    -1.51      0.133
## Year2011          -0.1175      0.0998    -1.18      0.240
## Year2012          -0.0799      0.1010    -0.79      0.429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0483, Adjusted R-squared:  0.0151
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 464 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.308  0.855   0.953   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      1.98e-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 1.23 1          1.109
## Year              1.23 16          1.006

```

Residuals from last author



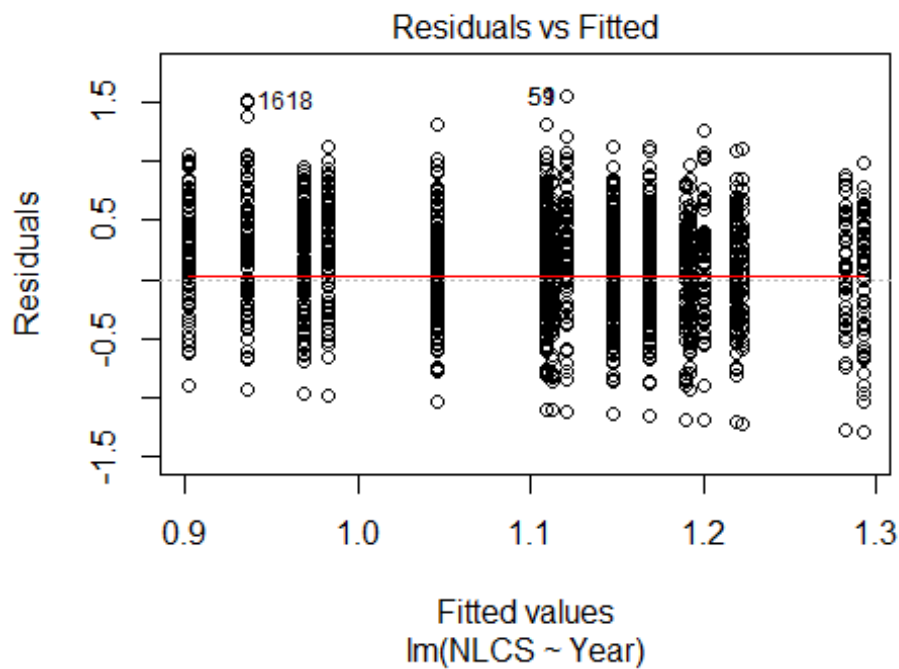
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2477 -0.2641 0.0179 0.2524 1.1753
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3103 0.0851 15.39 <2e-16 ***
## LastAuthorFemale1 0.0544 0.0483 1.13 0.260
## Year1997 -0.0510 0.1724 -0.30 0.768
## Year1998 0.0223 0.1524 0.15 0.884
## Year1999 -0.1494 0.2143 -0.70 0.486
## Year2000 -0.3954 0.1754 -2.25 0.025 *
## Year2001 -0.2643 0.1319 -2.00 0.046 *
## Year2002 -0.0454 0.1172 -0.39 0.699
## Year2003 -0.1740 0.1296 -1.34 0.180
## Year2004 -0.0368 0.1136 -0.32 0.746
## Year2005 -0.0323 0.1252 -0.26 0.796
## Year2006 -0.1850 0.1361 -1.36 0.175
```

```

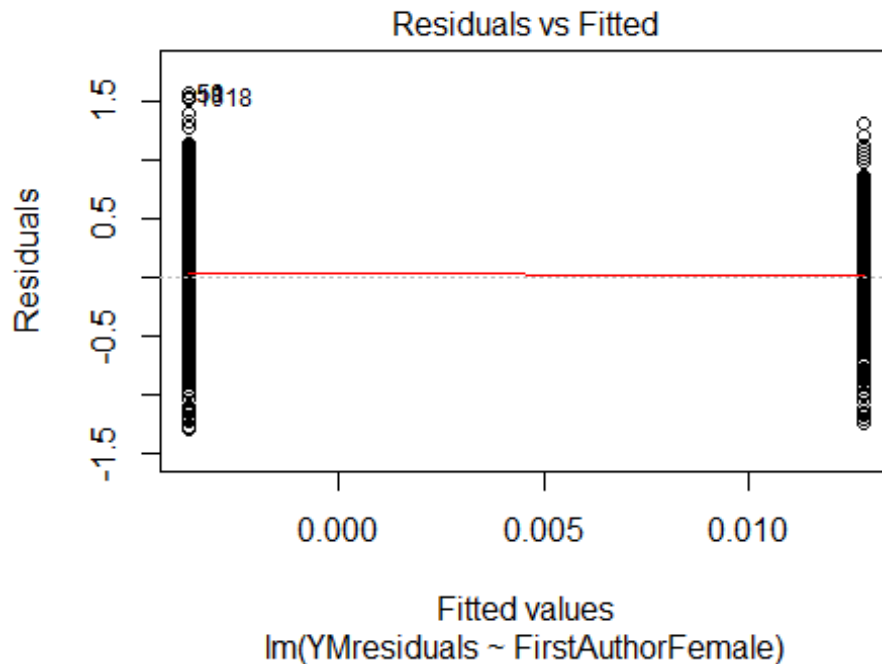
## Year2007          -0.2346      0.1098    -2.14      0.033 *
## Year2008          -0.1598      0.1085    -1.47      0.142
## Year2009          -0.1887      0.1004    -1.88      0.061 .
## Year2010          -0.1552      0.1174    -1.32      0.187
## Year2011          -0.0892      0.0973    -0.92      0.359
## Year2012          -0.0626      0.0997    -0.63      0.530
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0404, Adjusted R-squared:  0.00694
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 459 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.298  0.856  0.951   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      1.98e-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: 505"
## [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
## 172 137 175 165 151 221 215 200 153 172 206 225 225 336 347
## 2011 2012
## 388 381
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 45 58 63 51 81 120 129 93 98 115 120 114 160 165
## 2011 2012

```

```
## 191 194
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 72 39 52 60 46 77 108 118 82 90 93 110 96 140 147
## 2011 2012
## 159 158
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 80, df = 16, p-value = 2e-10
```

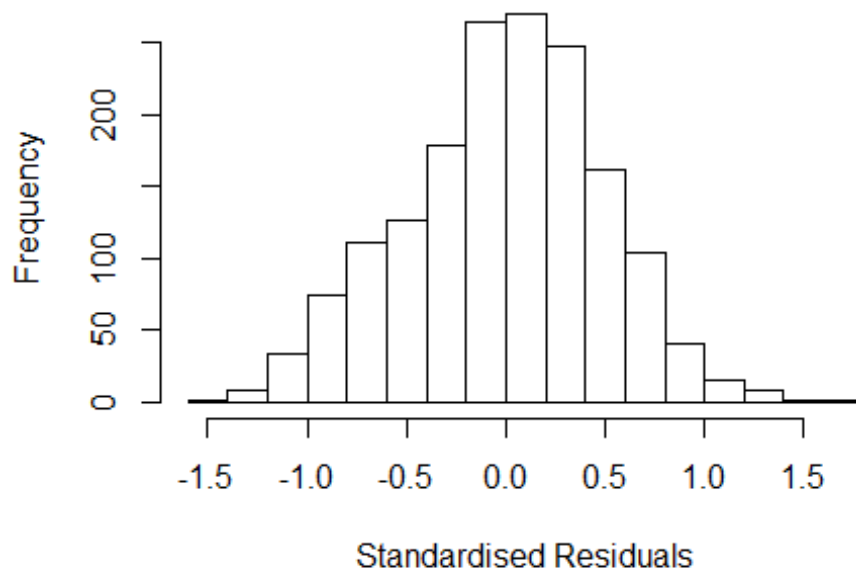


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.2, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 3.4813612629731"
## [1] "Male first author team size 2018 geometric mean: 3.10407102769895"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.24799405120277"
## [1] "Male last author team size 2018 geometric mean: 3.22226655173025"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, 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  1.079 1          1.039
## UniqueAuthors    1.288 4          1.032
## Year              1.350 16         1.009
```

Residuals from first and last author and team size



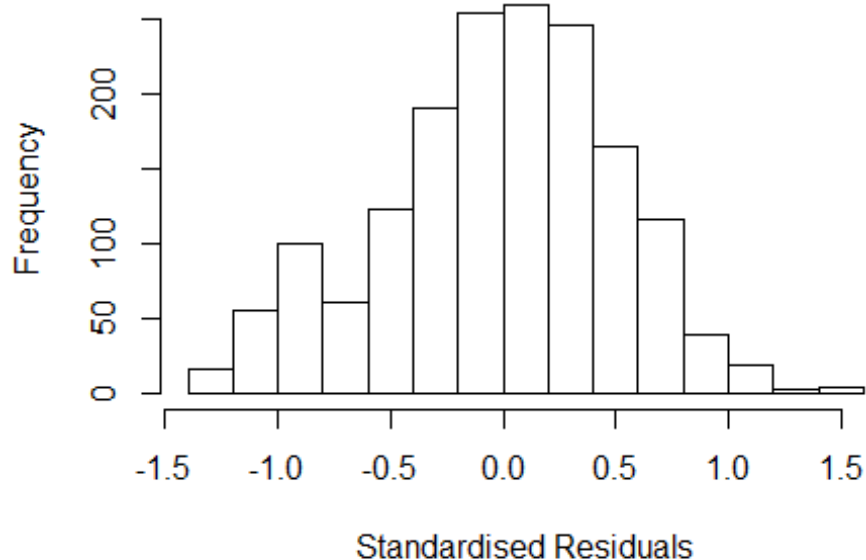
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4200 -0.3373 0.0229 0.3195 1.7965
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.897671 0.096271 9.32 < 2e-16 ***
## FirstAuthorFemale1 0.007757 0.029409 0.26 0.792
## LastAuthorFemale1 0.006406 0.033618 0.19 0.849
## UniqueAuthors2 0.297604 0.042120 7.07 2.4e-12 ***
## UniqueAuthors3 0.387860 0.041058 9.45 < 2e-16 ***
## UniqueAuthors4 0.300337 0.044468 6.75 2.0e-11 ***
## UniqueAuthors5 0.386478 0.048820 7.92 4.5e-15 ***
## Year1997 0.224692 0.122065 1.84 0.066 .
## Year1998 0.075836 0.116902 0.65 0.517
## Year1999 0.172905 0.117308 1.47 0.141
```

```

## Year2000          0.072999    0.114170    0.64    0.523
## Year2001          0.041713    0.105786    0.39    0.693
## Year2002         -0.234995    0.110016   -2.14    0.033 *
## Year2003         -0.245174    0.106029   -2.31    0.021 *
## Year2004         -0.207031    0.120307   -1.72    0.085 .
## Year2005          0.000101    0.100697    0.00    0.999
## Year2006          0.014038    0.099750    0.14    0.888
## Year2007         -0.171798    0.103815   -1.65    0.098 .
## Year2008          0.011123    0.100367    0.11    0.912
## Year2009         -0.014766    0.099863   -0.15    0.882
## Year2010         -0.112597    0.099747   -1.13    0.259
## Year2011         -0.044919    0.097646   -0.46    0.646
## Year2012         -0.060995    0.099742   -0.61    0.541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.123, Adjusted R-squared:  0.111
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 127 weights are ~= 1. The remaining 1520 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  0.856  0.951   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      6.07e-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 1.060 1      1.030
## LastAuthorFemale  1.054 1      1.027
## 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
## -1.3448 -0.3347  0.0218  0.3373  1.5678
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12229    0.10685   10.50  <2e-16 ***
## FirstAuthorFemale1  0.03952    0.02999    1.32   0.188
## LastAuthorFemale1 -0.00217    0.03458   -0.06   0.950
## Year1997         0.22252    0.13907    1.60   0.110
## Year1998         0.07469    0.13380    0.56   0.577
## Year1999         0.16815    0.13050    1.29   0.198
## Year2000         0.08284    0.12621    0.66   0.512
## Year2001         0.08770    0.12090    0.73   0.468
## Year2002        -0.22621    0.12494   -1.81   0.070 .
## Year2003        -0.24110    0.12155   -1.98   0.047 *
## Year2004        -0.17181    0.13688   -1.26   0.210
## Year2005         0.05508    0.11485    0.48   0.632
```

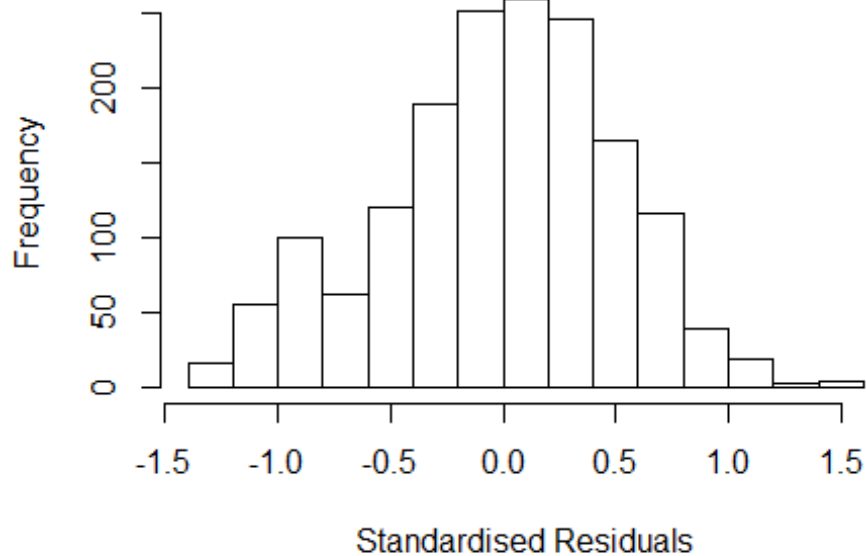


```

## Year2006          0.09064      0.11418      0.79      0.427
## Year2007         -0.13788      0.11995     -1.15      0.251
## Year2008          0.01751      0.11497      0.15      0.879
## Year2009          0.04528      0.11471      0.39      0.693
## Year2010         -0.07539      0.11478     -0.66      0.511
## Year2011          0.01300      0.11322      0.11      0.909
## Year2012          0.01324      0.11438      0.12      0.908
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.493
## Multiple R-squared:  0.0506, Adjusted R-squared:  0.0401
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1497 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.290  0.858   0.947   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      6.07e-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 1.04 1      1.020
## Year              1.04 16      1.001

```

Residuals from first author



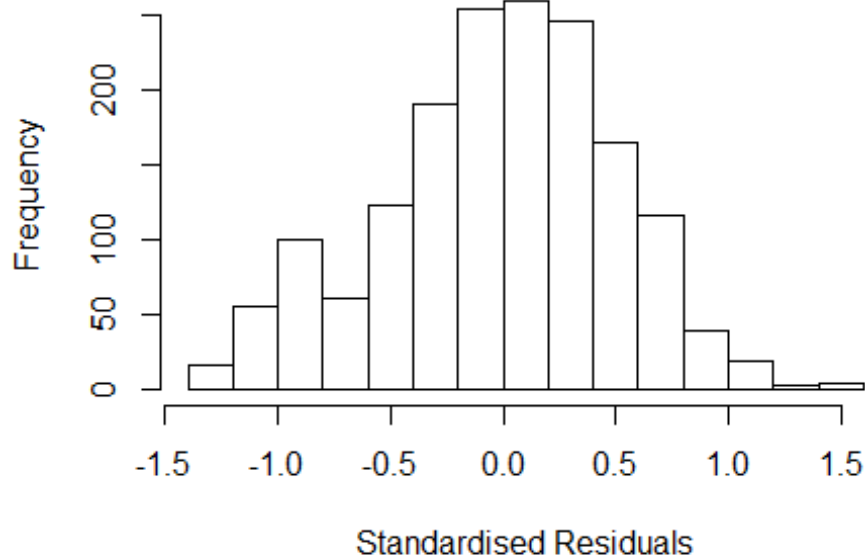
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3448 -0.3354  0.0213  0.3377  1.5680
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1224     0.1069   10.50  <2e-16 ***
## FirstAuthorFemale1  0.0391     0.0298    1.32   0.188
## Year1997          0.2224     0.1391    1.60   0.110
## Year1998          0.0745     0.1338    0.56   0.578
## Year1999          0.1678     0.1306    1.29   0.199
## Year2000          0.0827     0.1263    0.65   0.513
## Year2001          0.0875     0.1209    0.72   0.469
## Year2002         -0.2265     0.1250   -1.81   0.070 .
## Year2003         -0.2413     0.1216   -1.98   0.047 *
## Year2004         -0.1721     0.1370   -1.26   0.209
## Year2005          0.0547     0.1148    0.48   0.634
## Year2006          0.0904     0.1142    0.79   0.429
```

```

## Year2007          -0.1382      0.1199    -1.15     0.249
## Year2008           0.0171      0.1149     0.15     0.882
## Year2009           0.0450      0.1148     0.39     0.695
## Year2010          -0.0757      0.1148    -0.66     0.510
## Year2011           0.0125      0.1131     0.11     0.912
## Year2012           0.0128      0.1142     0.11     0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.492
## Multiple R-squared:  0.0507, Adjusted R-squared:  0.0407
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1498 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.858  0.947  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      6.07e-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 1.033 1          1.016
## Year              1.033 16          1.001

```

Residuals from last author



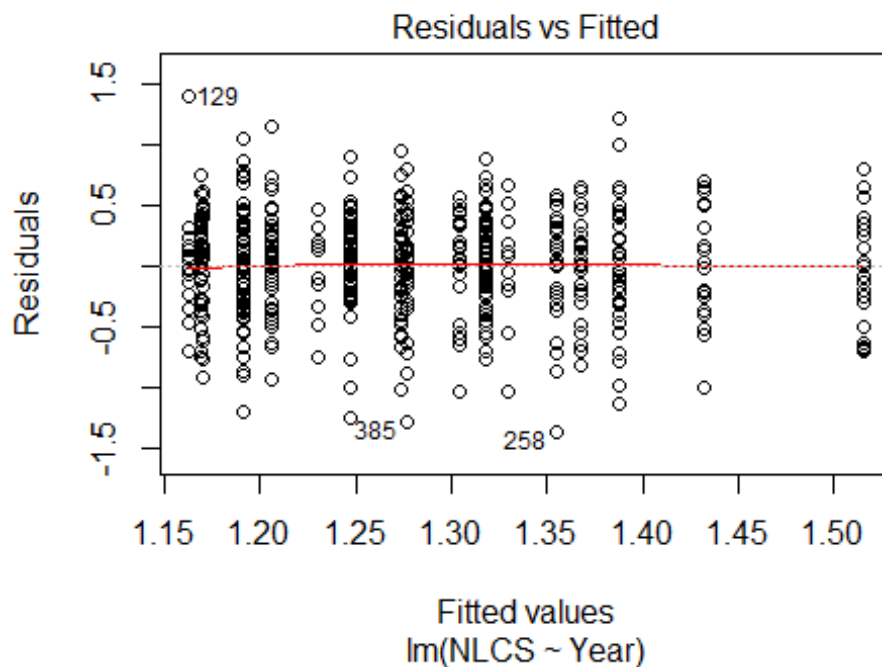
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3490 -0.3411  0.0212  0.3399  1.5580
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12497    0.10712   10.50  <2e-16 ***
## LastAuthorFemale1 0.00667    0.03424    0.19   0.846
## Year1997        0.22407    0.13953    1.61   0.109
## Year1998        0.07748    0.13421    0.58   0.564
## Year1999        0.17386    0.13042    1.33   0.183
## Year2000        0.08342    0.12613    0.66   0.508
## Year2001        0.09313    0.12125    0.77   0.443
## Year2002       -0.22322    0.12498   -1.79   0.074 .
## Year2003       -0.23399    0.12183   -1.92   0.055 .
## Year2004       -0.16823    0.13734   -1.22   0.221
## Year2005        0.05881    0.11523    0.51   0.610
## Year2006        0.09714    0.11447    0.85   0.396
```

```

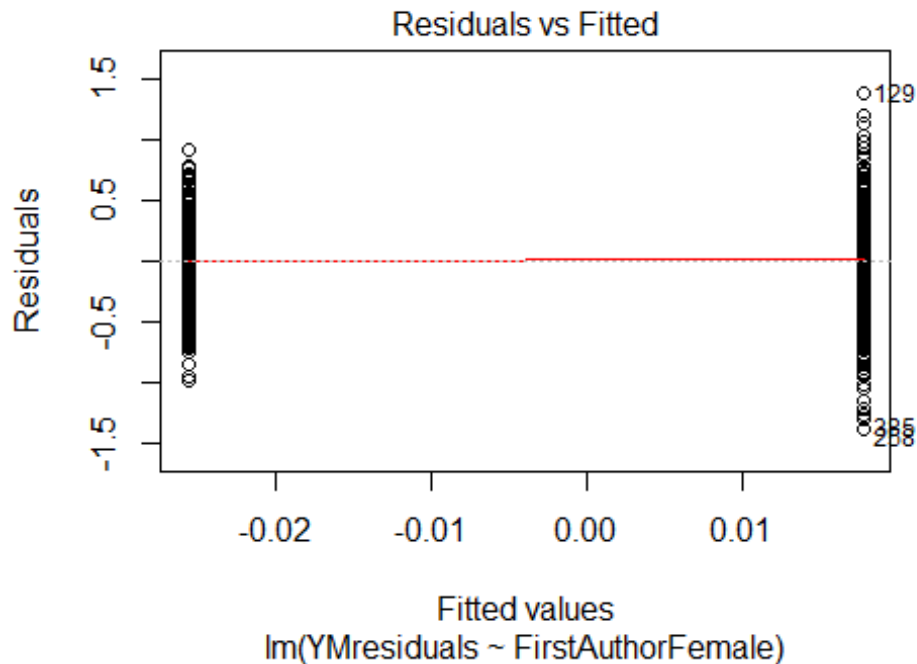
## Year2007          -0.13416      0.12027    -1.12      0.265
## Year2008           0.02395      0.11512      0.21      0.835
## Year2009           0.05189      0.11489      0.45      0.652
## Year2010          -0.06764      0.11508     -0.59      0.557
## Year2011           0.01882      0.11360      0.17      0.868
## Year2012           0.01889      0.11477      0.16      0.869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.492
## Multiple R-squared:  0.0497, Adjusted R-squared:  0.0398
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 156 weights are ~= 1. The remaining 1491 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.858  0.947  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      6.07e-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: 1647"
## [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
##   41   31   45   44   47   40   39   43   36   29   51   43   55   76   68
## 2011 2012
##   92   97
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21   11   19   25   25   11   22   32   25   21   31   34   38   58   46
## 2011 2012

```

```
## 68 62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 10 15 21 24 9 19 28 24 18 31 30 34 45 42
## 2011 2012
## 58 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 = 12, df = 16, p-value = 0.8
```



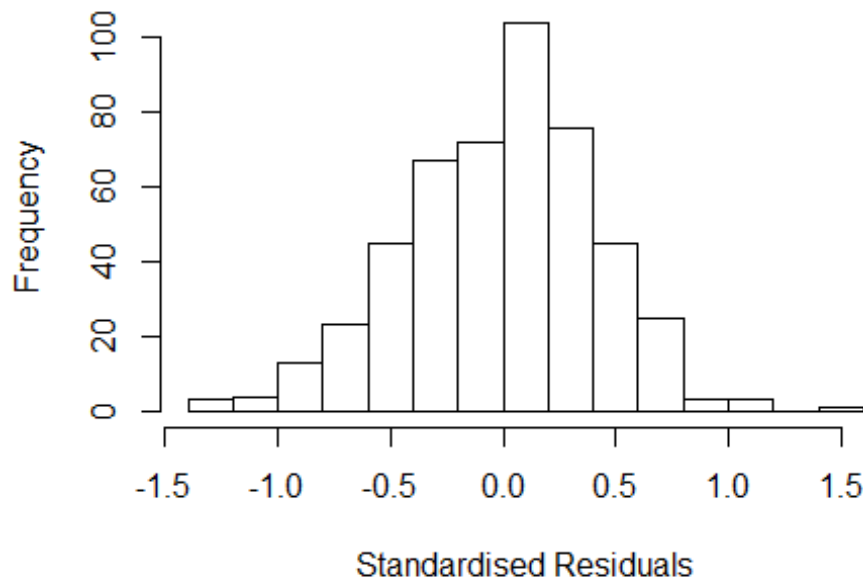
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 18, df = 1, p-value = 3e-05
## [1] "Female first author team size 2018 geometric mean: 4.83454939300973"
## [1] "Male first author team size 2018 geometric mean: 3.04872391452911"
## 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 = 1200, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.5432376714036"
## [1] "Male last author team size 2018 geometric mean: 4.00006396514882"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680, 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	1.148	1	1.072
LastAuthorFemale	1.285	1	1.133
UniqueAuthors	1.742	4	1.072
Year	2.143	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
## -1.3658 -0.3075 0.0311 0.2925 1.4245
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4304 0.1438 9.95 <2e-16 ***
## FirstAuthorFemale1 -0.0798 0.0401 -1.99 0.0472 *
## LastAuthorFemale1 0.0346 0.0486 0.71 0.4775
## UniqueAuthors2 0.1373 0.0937 1.47 0.1433
## UniqueAuthors3 0.1635 0.0893 1.83 0.0678 .
## UniqueAuthors4 0.1755 0.0954 1.84 0.0664 .
## UniqueAuthors5 0.1049 0.0822 1.28 0.2022
## Year1997 -0.1862 0.1965 -0.95 0.3438
## Year1998 -0.3959 0.1582 -2.50 0.0127 *
## Year1999 -0.1518 0.1560 -0.97 0.3312
```

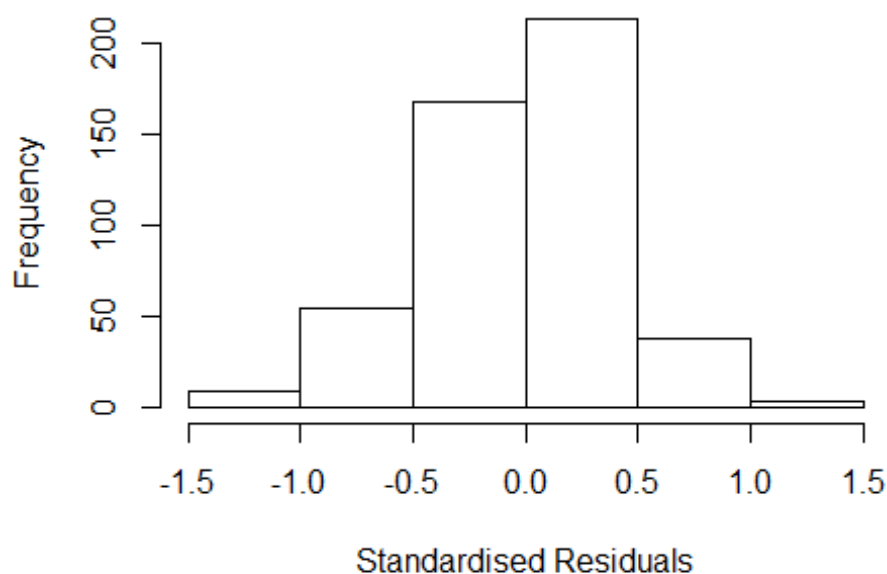


```

## Year2000          -0.1227      0.1615    -0.76    0.4477
## Year2001          -0.1832      0.1749    -1.05    0.2956
## Year2002          -0.0529      0.1683    -0.31    0.7535
## Year2003          -0.2390      0.1490    -1.60    0.1093
## Year2004          -0.2292      0.1540    -1.49    0.1375
## Year2005          -0.3171      0.1730    -1.83    0.0674 .
## Year2006          -0.1270      0.1638    -0.78    0.4386
## Year2007          -0.1951      0.1520    -1.28    0.2002
## Year2008          -0.3762      0.1410    -2.67    0.0079 **
## Year2009          -0.2552      0.1393    -1.83    0.0675 .
## Year2010          -0.3418      0.1405    -2.43    0.0153 *
## Year2011          -0.2281      0.1343    -1.70    0.0902 .
## Year2012          -0.3194      0.1417    -2.25    0.0246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.0693, Adjusted R-squared:  0.0249
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 442 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.218  0.873  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      2.07e-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 1.134 1      1.065
## LastAuthorFemale  1.187 1      1.090
## Year              1.314 16      1.009

```

Residuals from first and last author



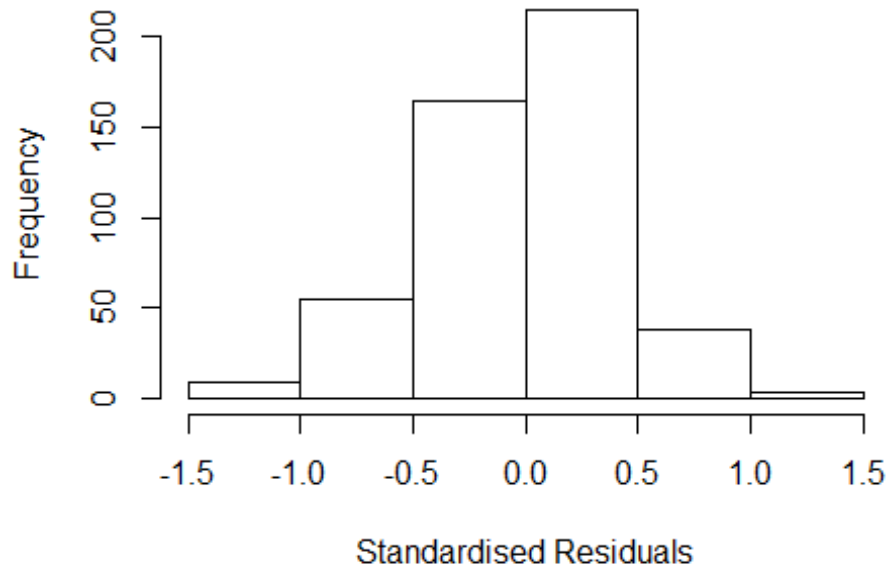
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4283 -0.2957 0.0236 0.3000 1.3985
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5551 0.1209 12.86 <2e-16 ***
## FirstAuthorFemale1 -0.0720 0.0406 -1.77 0.0771 .
## LastAuthorFemale1 0.0221 0.0477 0.46 0.6440
## Year1997 -0.1948 0.1976 -0.99 0.3248
## Year1998 -0.3896 0.1590 -2.45 0.0147 *
## Year1999 -0.1639 0.1536 -1.07 0.2863
## Year2000 -0.1268 0.1552 -0.82 0.4143
## Year2001 -0.2019 0.1657 -1.22 0.2237
## Year2002 -0.0463 0.1638 -0.28 0.7773
## Year2003 -0.2527 0.1448 -1.74 0.0817 .
## Year2004 -0.2132 0.1486 -1.43 0.1521
## Year2005 -0.3173 0.1710 -1.85 0.0643 .
```

```

## Year2006          -0.1308      0.1578   -0.83   0.4075
## Year2007          -0.2131      0.1482   -1.44   0.1513
## Year2008          -0.3776      0.1374   -2.75   0.0062 **
## Year2009          -0.2407      0.1353   -1.78   0.0757 .
## Year2010          -0.3498      0.1366   -2.56   0.0108 *
## Year2011          -0.2393      0.1299   -1.84   0.0661 .
## Year2012          -0.3241      0.1373   -2.36   0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0577, Adjusted R-squared:  0.0212
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.231  0.881  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      2.07e-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 1.118 1      1.057
## Year              1.118 16      1.003

```

Residuals from first author



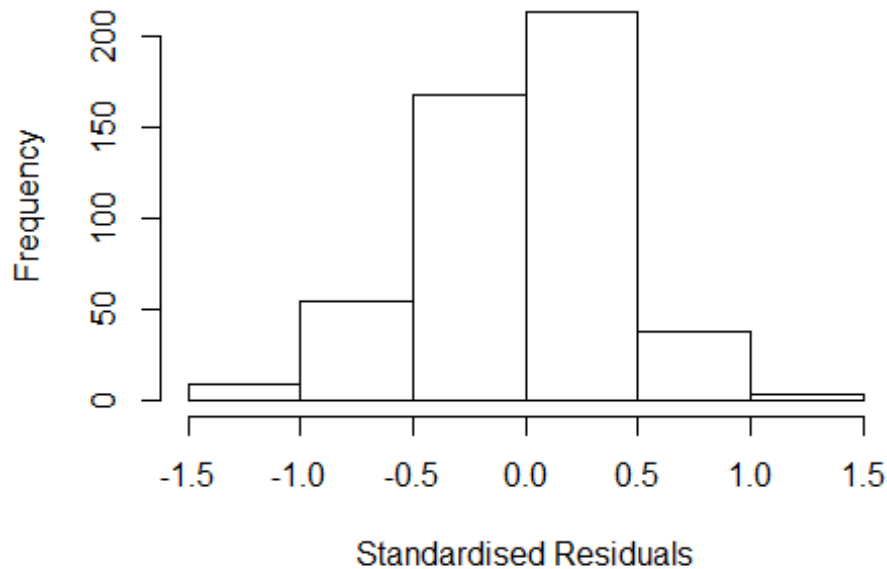
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4285 -0.2985 0.0237 0.2993 1.3928
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5567 0.1203 12.95 <2e-16 ***
## FirstAuthorFemale1 -0.0694 0.0404 -1.72 0.0865 .
## Year1997 -0.1881 0.1960 -0.96 0.3378
## Year1998 -0.3856 0.1589 -2.43 0.0156 *
## Year1999 -0.1627 0.1533 -1.06 0.2890
## Year2000 -0.1282 0.1546 -0.83 0.4073
## Year2001 -0.2022 0.1642 -1.23 0.2188
## Year2002 -0.0417 0.1619 -0.26 0.7969
## Year2003 -0.2487 0.1439 -1.73 0.0847 .
## Year2004 -0.2090 0.1475 -1.42 0.1573
## Year2005 -0.3191 0.1700 -1.88 0.0612 .
## Year2006 -0.1310 0.1572 -0.83 0.4050
```

```

## Year2007          -0.2103      0.1479   -1.42   0.1558
## Year2008          -0.3762      0.1367   -2.75   0.0062 **
## Year2009          -0.2378      0.1343   -1.77   0.0773 .
## Year2010          -0.3471      0.1360   -2.55   0.0110 *
## Year2011          -0.2360      0.1289   -1.83   0.0679 .
## Year2012          -0.3205      0.1363   -2.35   0.0192 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0574, Adjusted R-squared:  0.0231
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 438 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.880  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      2.07e-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 1.172 1          1.083
## Year            1.172 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
## -1.3961 -0.2941 0.0253 0.2952 1.4357
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5368 0.1189 12.92 <2e-16 ***
## LastAuthorFemale1 0.0110 0.0470 0.23 0.8154
## Year1997 -0.1938 0.1984 -0.98 0.3290
## Year1998 -0.4085 0.1562 -2.61 0.0092 **
## Year1999 -0.1676 0.1535 -1.09 0.2755
## Year2000 -0.1407 0.1551 -0.91 0.3645
## Year2001 -0.2016 0.1622 -1.24 0.2146
## Year2002 -0.0481 0.1632 -0.29 0.7684
## Year2003 -0.2590 0.1431 -1.81 0.0708 .
## Year2004 -0.2236 0.1480 -1.51 0.1314
## Year2005 -0.3263 0.1689 -1.93 0.0540 .
## Year2006 -0.1519 0.1552 -0.98 0.3284
```

```

## Year2007          -0.2333      0.1467    -1.59    0.1125
## Year2008          -0.3911      0.1357    -2.88    0.0041 **
## Year2009          -0.2485      0.1341    -1.85    0.0646 .
## Year2010          -0.3658      0.1346    -2.72    0.0068 **
## Year2011          -0.2554      0.1279    -2.00    0.0463 *
## Year2012          -0.3247      0.1355    -2.40    0.0170 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0507, Adjusted R-squared:  0.0161
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 435 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.248  0.878  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      2.07e-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: 484"
## [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]"
##
## 2010 2012
##    1    2
##
## 2010 2012
##    1    1
##
## 2010 2012
##    1    1

```

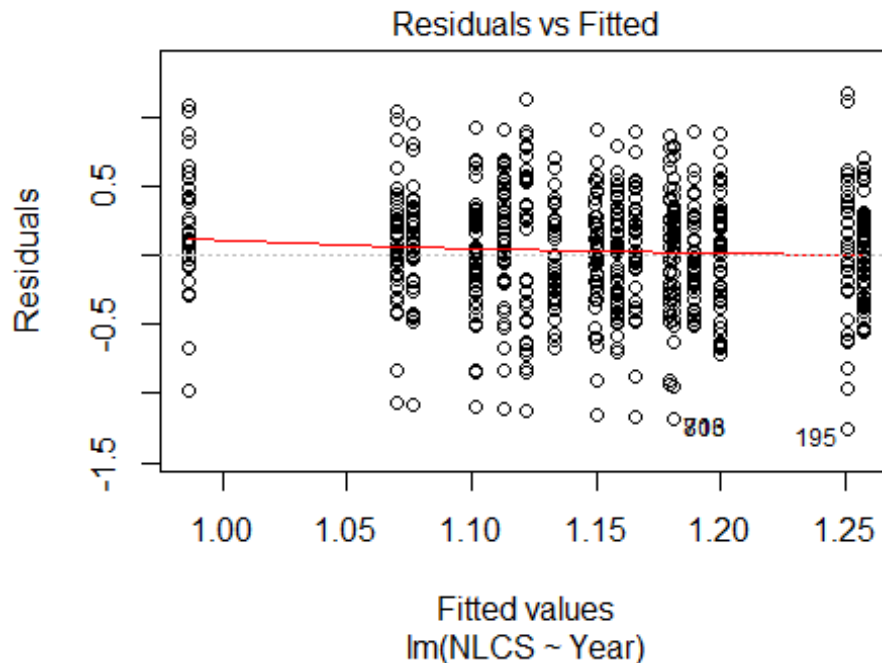
```

## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 8"
## [1] "Male first author team size 2018 geometric mean: 4"
##
## Wilcoxon rank sum test
##
## 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: 4"
## [1] "Male last author team size 2018 geometric mean: 8"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, 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"
## [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 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
## 86 100 99 89 81 73 89 76 74 85 99 95 103 104 85
## 2011 2012
## 90 92
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 32 33 40 33 15 51 42 37 47 48 42 53 54 43
## 2011 2012
## 58 52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 29 30 31 27 12 45 37 35 38 41 39 51 48 41
## 2011 2012
## 56 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 = 60, df = 16, p-value = 6e-07
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.043, df = 1, p-value = 0.8

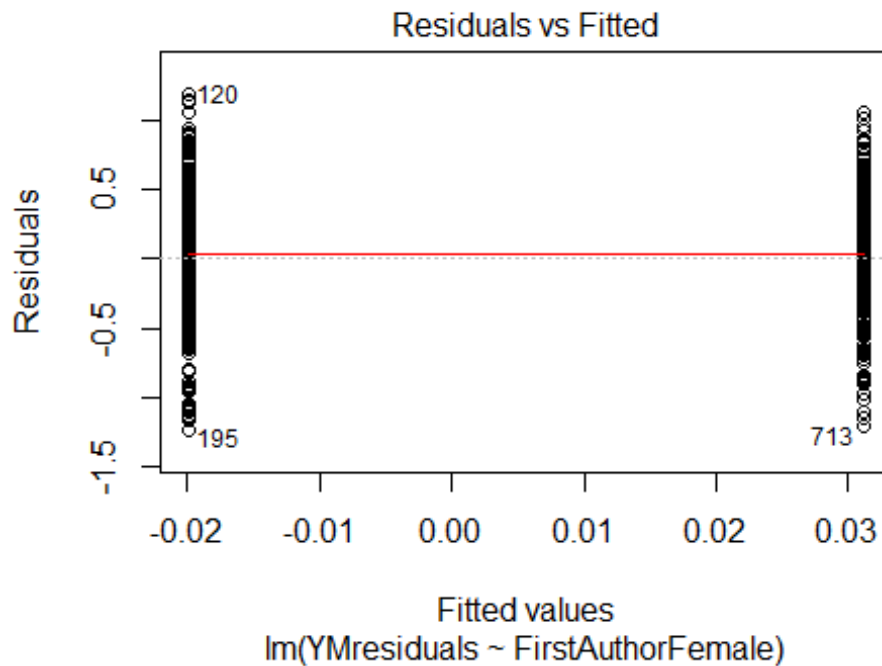
## [1] "Female first author team size 2018 geometric mean: 4.1819227431901"
## [1] "Male first author team size 2018 geometric mean: 3.31523446472126"

## 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 = 370, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
```

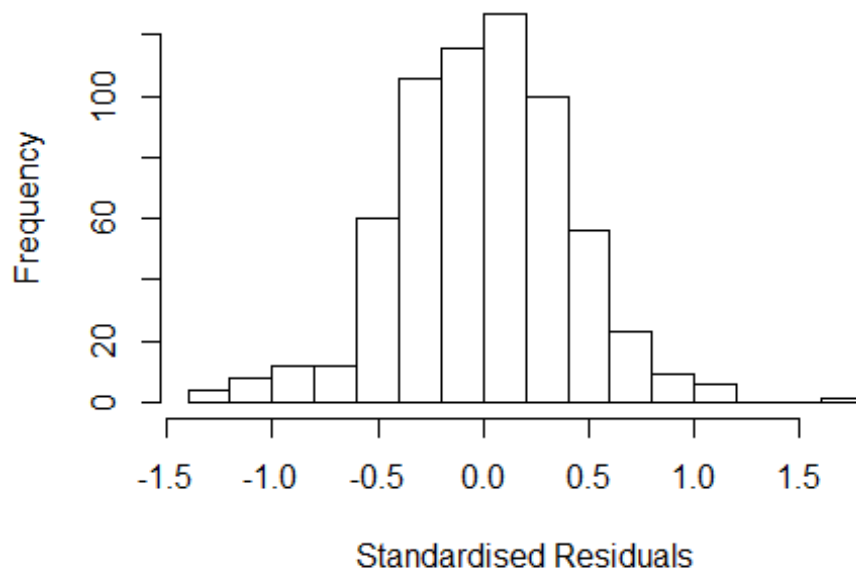
```
## [1] "Female last author team size 2018 geometric mean: 3.0861636884341"
## [1] "Male last author team size 2018 geometric mean: 3.95084719768429"

## 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 1.210  1      1.100
## LastAuthorFemale  1.217  1      1.103
## UniqueAuthors    1.776  4      1.074
## Year             2.233 16      1.025
```

Residuals from first and last author and team size



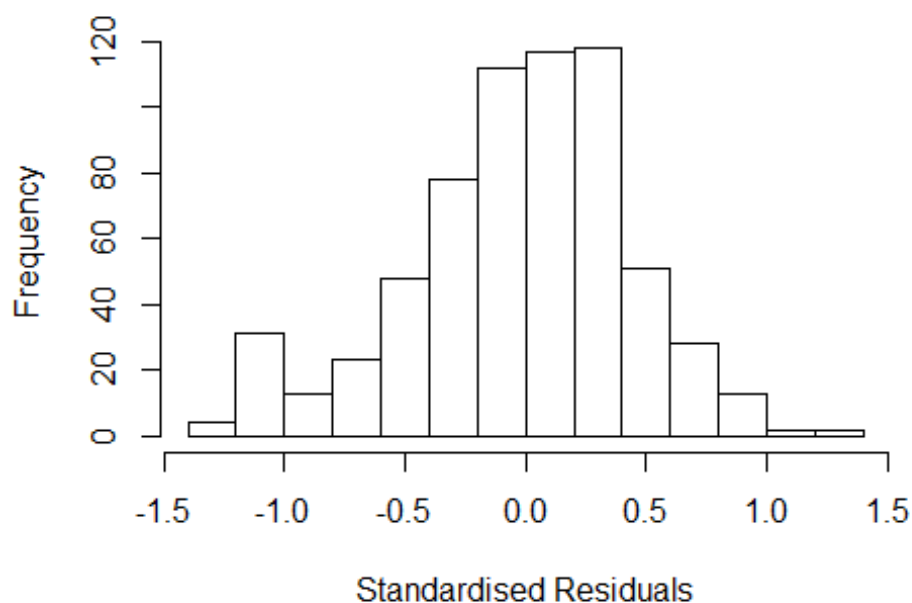
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35460 -0.29208  0.00863  0.26012  1.70181
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.42138    0.12934   3.26   0.0012 **
## FirstAuthorFemale1 0.01031    0.03584   0.29   0.7737
## LastAuthorFemale1 0.03220    0.04342   0.74   0.4586
## UniqueAuthors2    0.73282    0.09078   8.07 3.6e-15 ***
## UniqueAuthors3    0.78543    0.08834   8.89 < 2e-16 ***
## UniqueAuthors4    0.79623    0.09096   8.75 < 2e-16 ***
## UniqueAuthors5    0.92895    0.08651  10.74 < 2e-16 ***
## Year1997          0.04215    0.15591   0.27   0.7870
## Year1998          0.12164    0.14559   0.84   0.4038
## Year1999         -0.10231    0.14963  -0.68   0.4944
```

```

## Year2000      -0.00703    0.15261   -0.05    0.9633
## Year2001      -0.00115    0.18932   -0.01    0.9952
## Year2002      -0.11169    0.13676   -0.82    0.4144
## Year2003      -0.00603    0.14532   -0.04    0.9669
## Year2004      -0.12217    0.14681   -0.83    0.4056
## Year2005      -0.02453    0.14216   -0.17    0.8631
## Year2006       0.02195    0.14385    0.15    0.8788
## Year2007      -0.08142    0.13748   -0.59    0.5539
## Year2008      -0.11670    0.13411   -0.87    0.3845
## Year2009      -0.00694    0.13682   -0.05    0.9596
## Year2010      -0.01316    0.14112   -0.09    0.9257
## Year2011      -0.06438    0.13408   -0.48    0.6313
## Year2012      -0.09038    0.13789   -0.66    0.5124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.28,   Adjusted R-squared:  0.254
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.010  0.876  0.944  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      1.56e-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 1.17 1      1.082
## LastAuthorFemale  1.17 1      1.081
## Year              1.36 16      1.010

```

Residuals from first and last author



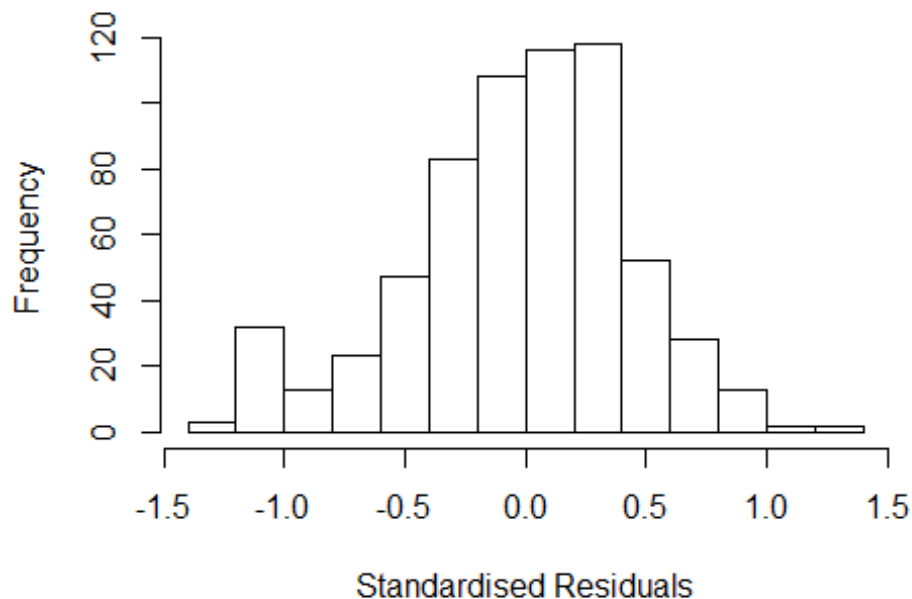
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2473 -0.2993  0.0157  0.2743  1.2283
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03314    0.16710   6.18 1.1e-09 ***
## FirstAuthorFemale1  0.04061    0.03951   1.03  0.30
## LastAuthorFemale1  0.01949    0.05135   0.38  0.70
## Year1997         0.15654    0.19651   0.80  0.43
## Year1998         0.13630    0.18249   0.75  0.46
## Year1999         0.00393    0.18581   0.02  0.98
## Year2000         0.15271    0.18101   0.84  0.40
## Year2001         0.05913    0.19903   0.30  0.77
## Year2002         0.03676    0.17992   0.20  0.84
## Year2003         0.17356    0.18205   0.95  0.34
## Year2004         0.08448    0.17624   0.48  0.63
## Year2005         0.09906    0.17933   0.55  0.58
```

```

## Year2006          0.20108      0.17877      1.12      0.26
## Year2007          0.13565      0.17218      0.79      0.43
## Year2008          0.04002      0.17356      0.23      0.82
## Year2009          0.17734      0.17126      1.04      0.30
## Year2010          0.14596      0.17917      0.81      0.42
## Year2011          0.14280      0.18021      0.79      0.43
## Year2012          0.01741      0.21057      0.08      0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0227, Adjusted R-squared:  -0.00567
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 587 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.348  0.869  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      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 1.18 1      1.086
## Year              1.18 16      1.005

```

Residuals from first author



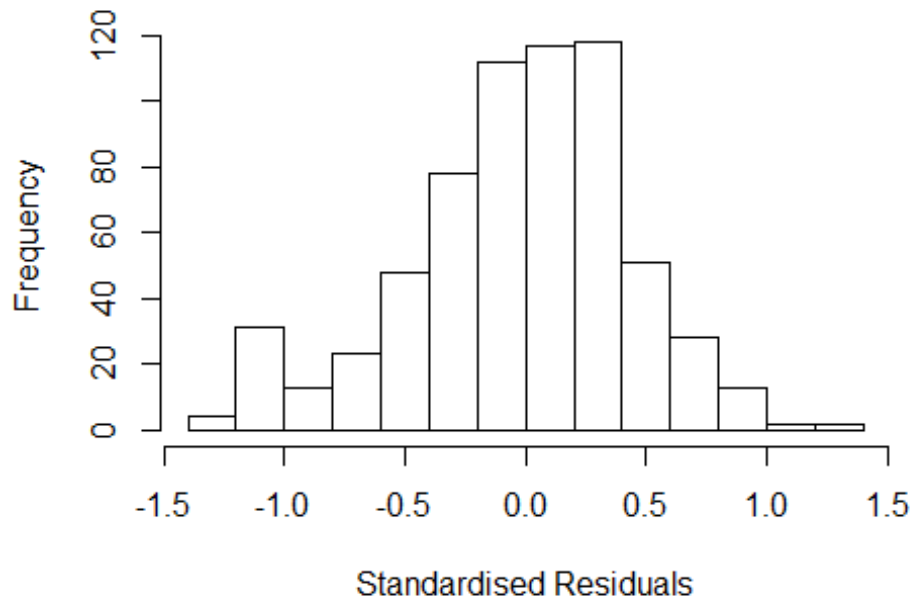
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2503 -0.2889  0.0154  0.2731  1.2256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03773    0.16496   6.29   6e-10 ***
## FirstAuthorFemale1 0.04204    0.04005   1.05    0.29
## Year1997        0.15471    0.19573   0.79    0.43
## Year1998        0.13476    0.18152   0.74    0.46
## Year1999        0.00111    0.18438   0.01    1.00
## Year2000        0.14894    0.17937   0.83    0.41
## Year2001        0.05516    0.19791   0.28    0.78
## Year2002        0.03398    0.17849   0.19    0.85
## Year2003        0.17049    0.18076   0.94    0.35
## Year2004        0.08158    0.17480   0.47    0.64
## Year2005        0.09757    0.17814   0.55    0.58
## Year2006        0.19940    0.17745   1.12    0.26
```

```

## Year2007          0.13751    0.17101    0.80    0.42
## Year2008          0.03937    0.17248    0.23    0.82
## Year2009          0.17569    0.16986    1.03    0.30
## Year2010          0.14463    0.17811    0.81    0.42
## Year2011          0.14204    0.17915    0.79    0.43
## Year2012          0.01454    0.20849    0.07    0.94
## ---
## 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:  -0.00425
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 587 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.346  0.868  0.952  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      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 1.18 1          1.086
## Year            1.18 16          1.005

```


Residuals from last author



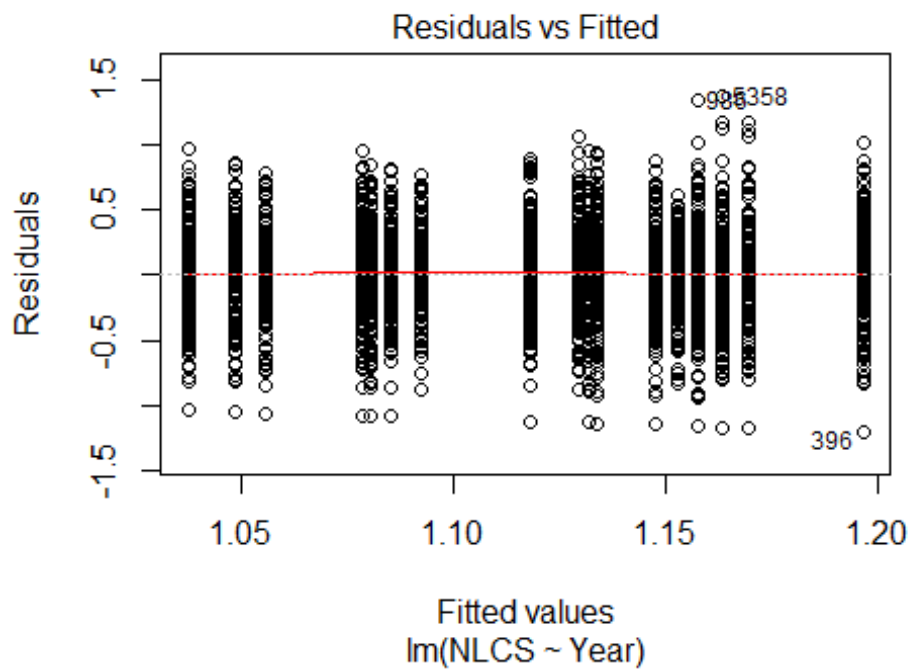
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2248 -0.2864  0.0107  0.2792  1.2175
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05157    0.16152   6.51 1.5e-10 ***
## LastAuthorFemale1 0.02424    0.05239   0.46  0.64
## Year1997        0.14896    0.19271   0.77  0.44
## Year1998        0.13560    0.17911   0.76  0.45
## Year1999       -0.00171    0.18226  -0.01  0.99
## Year2000        0.15684    0.17787   0.88  0.38
## Year2001        0.05400    0.19603   0.28  0.78
## Year2002        0.02843    0.17577   0.16  0.87
## Year2003        0.17146    0.17790   0.96  0.34
## Year2004        0.08062    0.17246   0.47  0.64
## Year2005        0.09213    0.17513   0.53  0.60
## Year2006        0.19479    0.17501   1.11  0.27
```

```

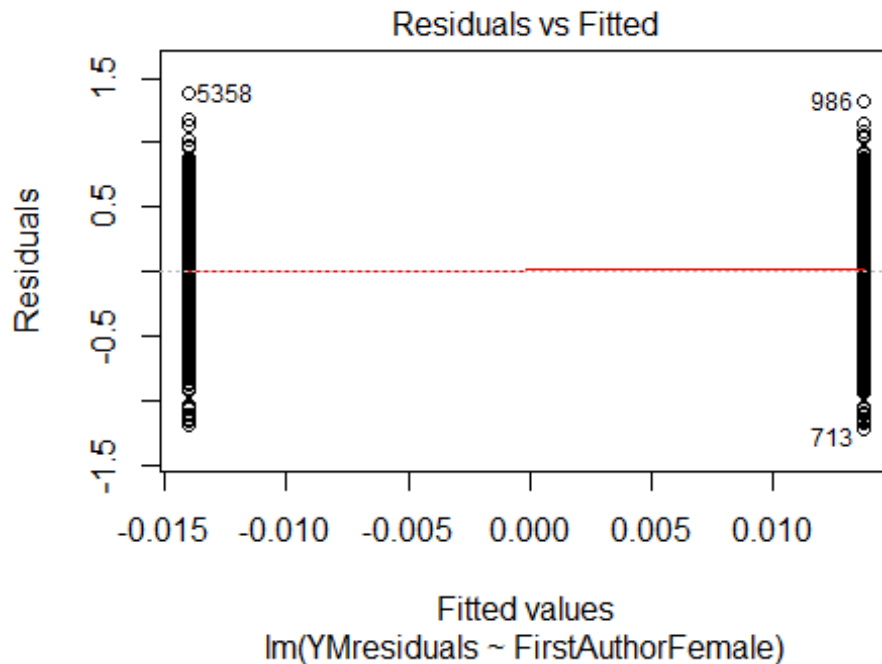
## Year2007      0.13457      0.16855      0.80      0.42
## Year2008      0.04216      0.16980      0.25      0.80
## Year2009      0.17490      0.16759      1.04      0.30
## Year2010      0.14334      0.17560      0.82      0.41
## Year2011      0.13527      0.17636      0.77      0.44
## Year2012      0.00575      0.20602      0.03      0.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.0211, Adjusted R-squared:  -0.00561
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 589 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.368  0.867  0.951  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      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: 640"
## [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
## 289 319 337 297 351 371 353 243 273 258 324 317 324 268 244
## 2011 2012
## 262 325
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 163 182 192 182 153 107 227 163 162 172 216 215 234 181 180
## 2011 2012

```

```
## 182 217
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 148 160 169 157 135 94 199 132 139 145 195 190 206 165 159
## 2011 2012
## 161 192
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 65, df = 16, p-value = 7e-08
```

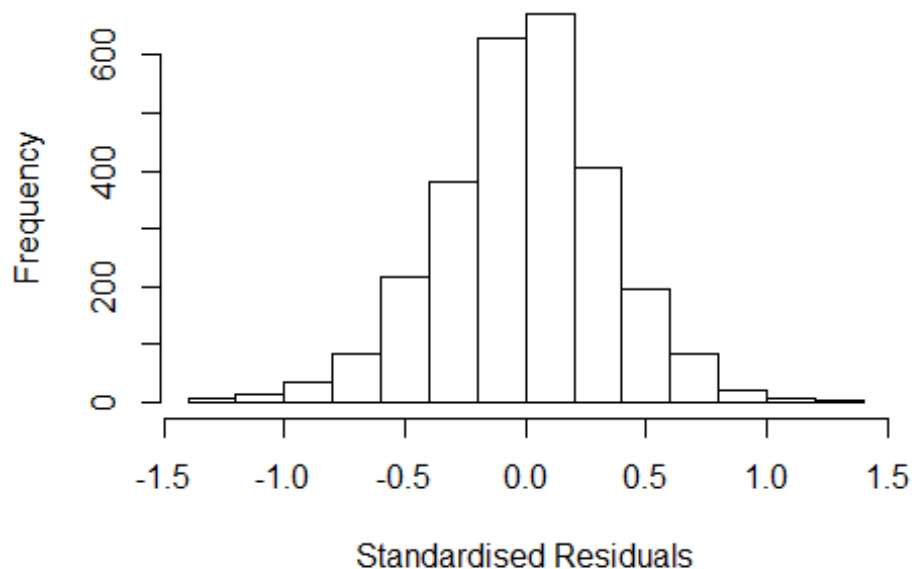


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.3, df = 1, p-value = 0.007
```



```
## [1] "Female first author team size 2018 geometric mean: 6.73747939301847"
## [1] "Male first author team size 2018 geometric mean: 5.53781129019183"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.49245745867671"
## [1] "Male last author team size 2018 geometric mean: 6.08078830989859"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3600, 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 1.044 1      1.022
## LastAuthorFemale  1.044 1      1.022
## UniqueAuthors    1.244 4      1.028
## Year             1.273 16      1.008
```

Residuals from first and last author and team size



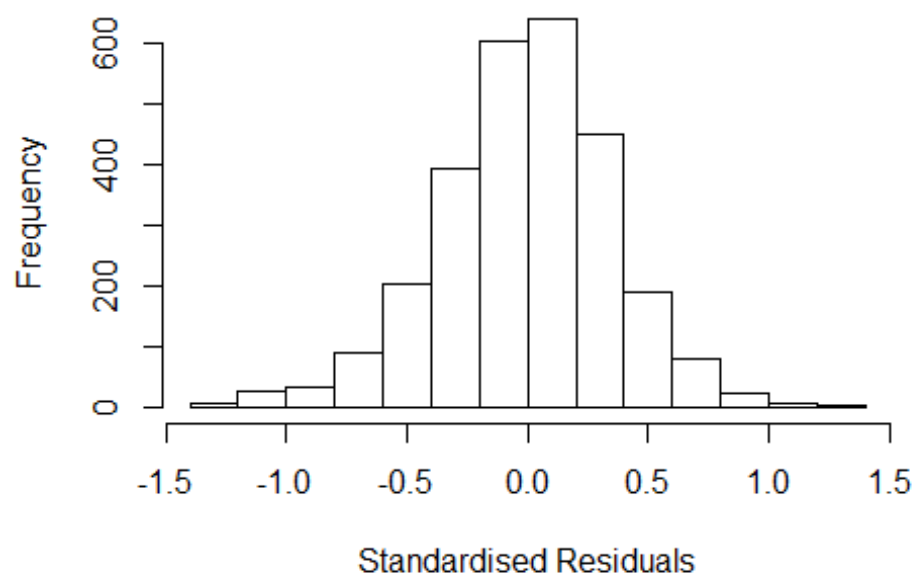
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31204 -0.21589  0.00521  0.20982  1.33562
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9988    0.0655   15.25 < 2e-16 ***
## FirstAuthorFemale1 -0.0369    0.0134   -2.76  0.0058 **
## LastAuthorFemale1 -0.0366    0.0158   -2.32  0.0205 *
## UniqueAuthors2     0.1025    0.0636    1.61  0.1071
## UniqueAuthors3     0.1459    0.0613    2.38  0.0173 *
## UniqueAuthors4     0.1885    0.0612    3.08  0.0021 **
## UniqueAuthors5     0.2888    0.0604    4.78 1.8e-06 ***
## Year1997          0.0570    0.0442    1.29  0.1973
## Year1998         -0.0198    0.0420   -0.47  0.6381
## Year1999         -0.0829    0.0381   -2.17  0.0298 *
```

```

## Year2000          0.0245      0.0508      0.48      0.6292
## Year2001         -0.0622      0.0494     -1.26      0.2080
## Year2002         -0.0294      0.0375     -0.78      0.4328
## Year2003         -0.1205      0.0443     -2.72      0.0066 **
## Year2004         -0.1332      0.0435     -3.06      0.0022 **
## Year2005         -0.0388      0.0387     -1.00      0.3169
## Year2006         -0.1140      0.0382     -2.98      0.0029 **
## Year2007         -0.0976      0.0373     -2.62      0.0089 **
## Year2008         -0.1257      0.0388     -3.23      0.0012 **
## Year2009         -0.0563      0.0416     -1.35      0.1757
## Year2010         -0.0688      0.0388     -1.77      0.0766 .
## Year2011         -0.0803      0.0423     -1.90      0.0574 .
## Year2012         -0.0825      0.0410     -2.01      0.0441 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.323
## Multiple R-squared:  0.0766, Adjusted R-squared:  0.0691
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 225 weights are ~= 1. The remaining 2521 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0483 0.8590 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      3.64e-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 1.034 1      1.017
## LastAuthorFemale  1.034 1      1.017
## Year              1.062 16      1.002

```

Residuals from first and last author



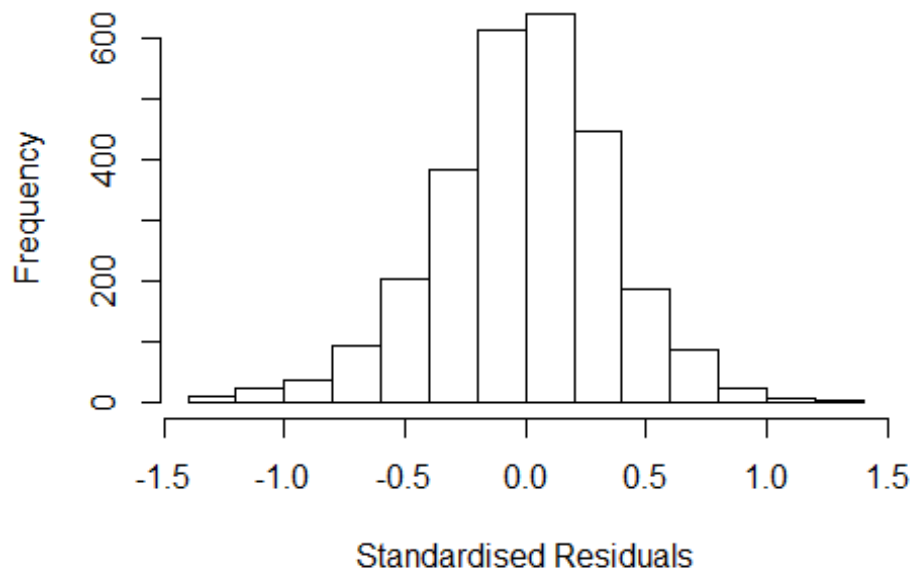
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24917 -0.22432  0.00585  0.22492  1.37909
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.184113   0.029629   39.96  <2e-16 ***
## FirstAuthorFemale1 -0.032470   0.013615   -2.38   0.0172 *
## LastAuthorFemale1 -0.046040   0.016209   -2.84   0.0045 **
## Year1997         0.065059   0.044861    1.45   0.1471
## Year1998        -0.010982   0.040697   -0.27   0.7873
## Year1999        -0.080461   0.036644   -2.20   0.0282 *
## Year2000         0.028037   0.049155    0.57   0.5685
## Year2001        -0.040048   0.050210   -0.80   0.4252
## Year2002         0.004110   0.036659    0.11   0.9108
## Year2003        -0.090093   0.043676   -2.06   0.0392 *
## Year2004        -0.096262   0.044138   -2.18   0.0293 *
## Year2005         0.005832   0.037723    0.15   0.8772
```

```

## Year2006      -0.071463    0.037956   -1.88    0.0598 .
## Year2007      -0.058434    0.036037   -1.62    0.1050
## Year2008      -0.077741    0.037292   -2.08    0.0372 *
## Year2009      -0.000734    0.041355   -0.02    0.9858
## Year2010      -0.012601    0.037367   -0.34    0.7360
## Year2011      -0.042493    0.040877   -1.04    0.2987
## Year2012      -0.026217    0.039806   -0.66    0.5102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.016
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 274 weights are ~= 1. The remaining 2472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.038  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      3.64e-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 1.031 1      1.015
## Year      1.031 16      1.001

```


Residuals from first author



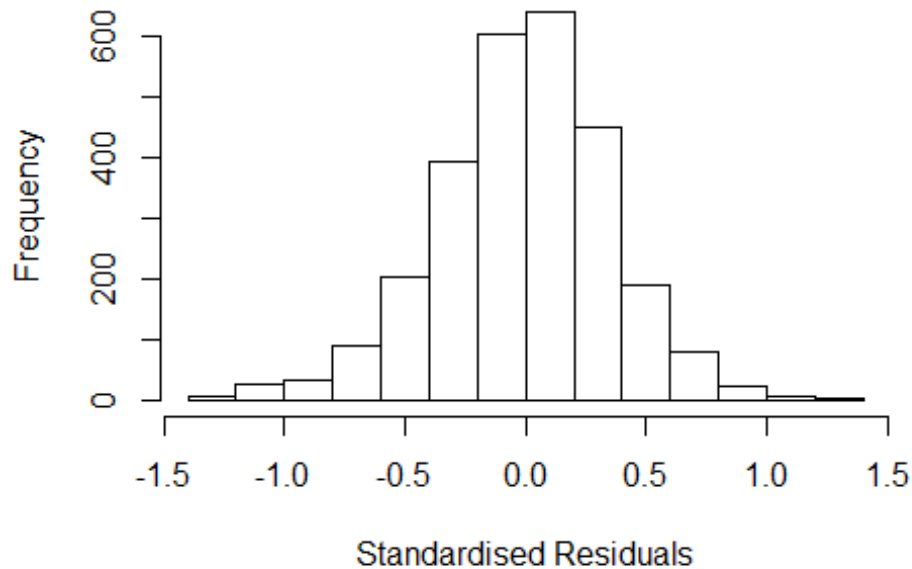
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24243 -0.22359  0.00403  0.22470  1.39437
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.17750    0.02953   39.87  <2e-16 ***
## FirstAuthorFemale1 -0.03516    0.01362   -2.58  0.0099 **
## Year1997          0.06492    0.04512    1.44  0.1503
## Year1998         -0.01008    0.04071   -0.25  0.8045
## Year1999         -0.08260    0.03677   -2.25  0.0248 *
## Year2000          0.03154    0.04902    0.64  0.5200
## Year2001         -0.04306    0.05051   -0.85  0.3940
## Year2002          0.00258    0.03668    0.07  0.9439
## Year2003         -0.09184    0.04366   -2.10  0.0355 *
## Year2004         -0.09839    0.04393   -2.24  0.0252 *
## Year2005          0.00303    0.03782    0.08  0.9361
## Year2006         -0.07281    0.03786   -1.92  0.0545 .
```

```

## Year2007          -0.06081    0.03614   -1.68    0.0926 .
## Year2008          -0.08102    0.03734   -2.17    0.0301 *
## Year2009          -0.00672    0.04126   -0.16    0.8707
## Year2010          -0.01728    0.03725   -0.46    0.6427
## Year2011          -0.04760    0.04108   -1.16    0.2466
## Year2012          -0.03247    0.03962   -0.82    0.4126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0195, Adjusted R-squared:  0.0133
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 266 weights are ~= 1. The remaining 2480 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0317 0.8600 0.9480 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      3.64e-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 1.032 1          1.016
## Year            1.032 16          1.001

```

Residuals from last author



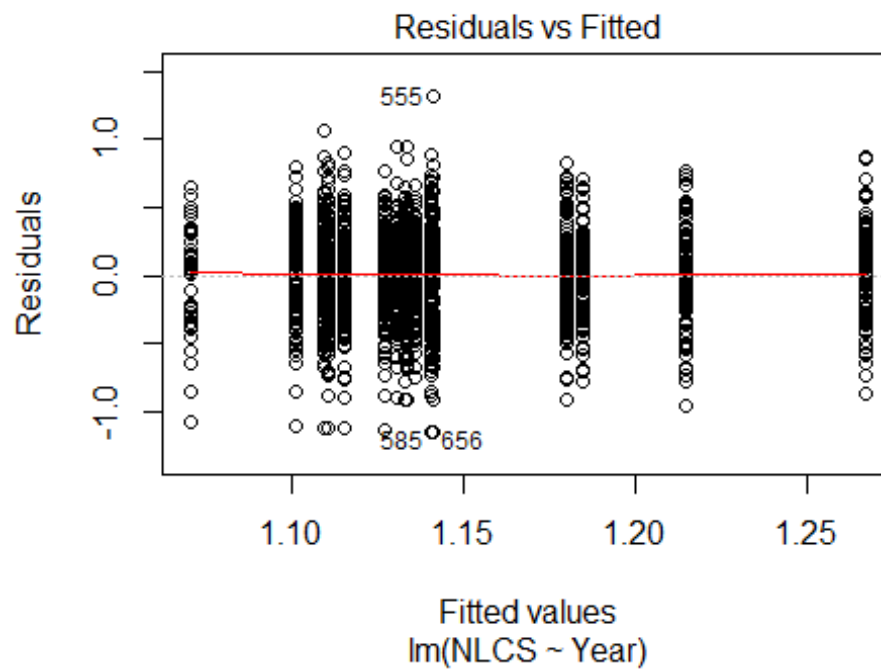
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23522 -0.22475  0.00547  0.22213  1.36294
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17120    0.02913   40.21  <2e-16 ***
## LastAuthorFemale1 -0.04874    0.01617   -3.01  0.0026 **
## Year1997         0.06401    0.04511    1.42  0.1560
## Year1998        -0.00957    0.04098   -0.23  0.8153
## Year1999        -0.08082    0.03688   -2.19  0.0285 *
## Year2000         0.02563    0.04943    0.52  0.6042
## Year2001        -0.04217    0.05069   -0.83  0.4055
## Year2002         0.00165    0.03681    0.04  0.9642
## Year2003        -0.09286    0.04397   -2.11  0.0348 *
## Year2004        -0.10273    0.04399   -2.34  0.0196 *
## Year2005         0.00202    0.03789    0.05  0.9575
## Year2006        -0.07485    0.03806   -1.97  0.0494 *
```

```

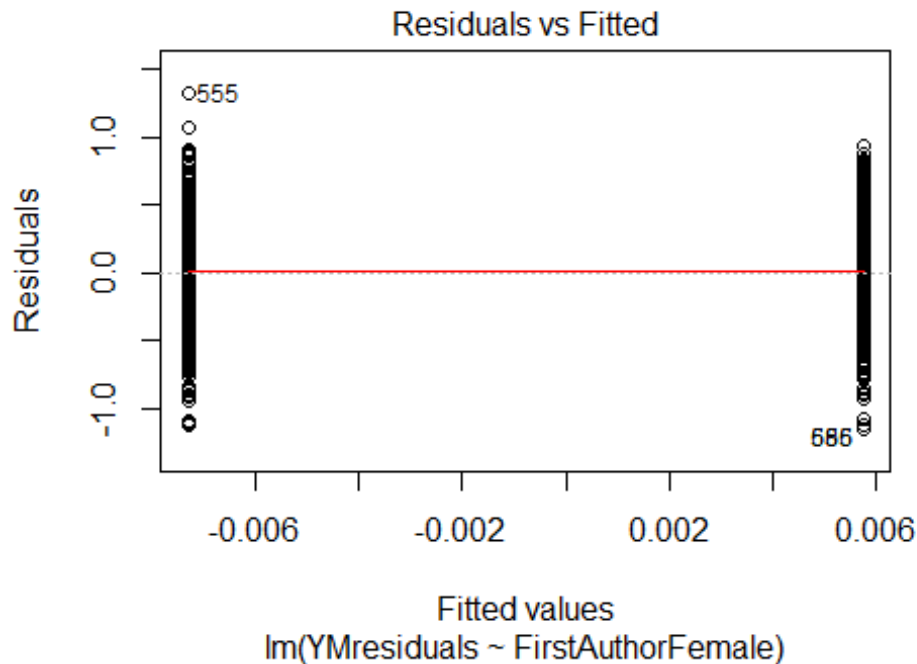
## Year2007          -0.06332      0.03609      -1.75      0.0794 .
## Year2008          -0.08134      0.03740      -2.18      0.0297 *
## Year2009          -0.00414      0.04146      -0.10      0.9204
## Year2010          -0.01720      0.03758      -0.46      0.6472
## Year2011          -0.04593      0.04094      -1.12      0.2621
## Year2012          -0.03102      0.03994      -0.78      0.4375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0202, Adjusted R-squared:  0.0141
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 266 weights are ~= 1. The remaining 2480 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0458 0.8570 0.9480 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      3.64e-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: 2746"
## [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
## 178 211 168 163 160 165 209 159 158 168 163 174 170 156 155
## 2011 2012
## 167 146
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 113 71 91 64 42 140 109 108 110 101 112 120 107 113
## 2011 2012

```

```
## 134 105
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 101 65 81 55 38 129 89 95 98 88 100 110 98 101
## 2011 2012
## 121 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 = 26, df = 16, p-value = 0.06
```

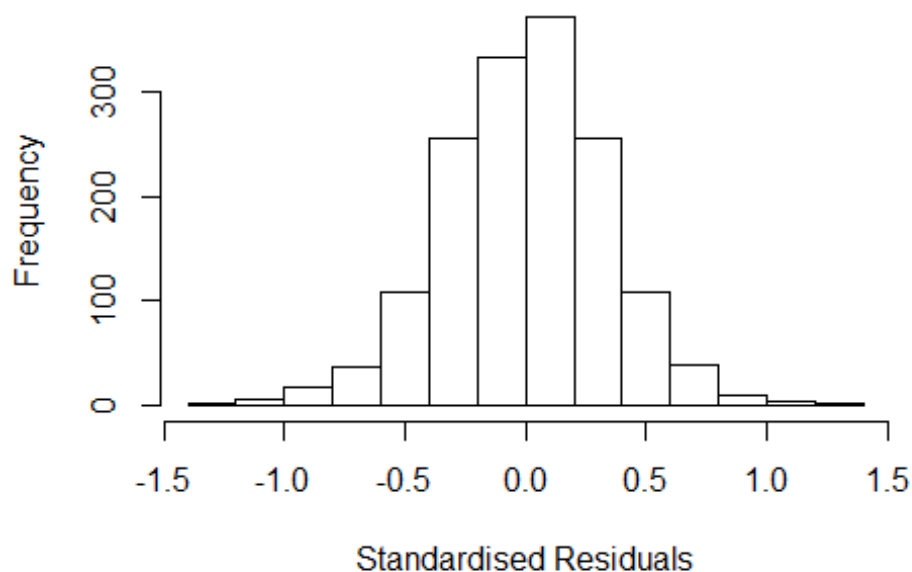


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.75, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 5.32436037326563"
## [1] "Male first author team size 2018 geometric mean: 3.65424028549649"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3300, p-value = 0.004
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.84479334105386"
## [1] "Male last author team size 2018 geometric mean: 4.4952807011935"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, 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 1.060 1          1.029
## LastAuthorFemale  1.062 1          1.031
## UniqueAuthors    1.204 4          1.024
## Year             1.292 16          1.008
```

Residuals from first and last author and team size



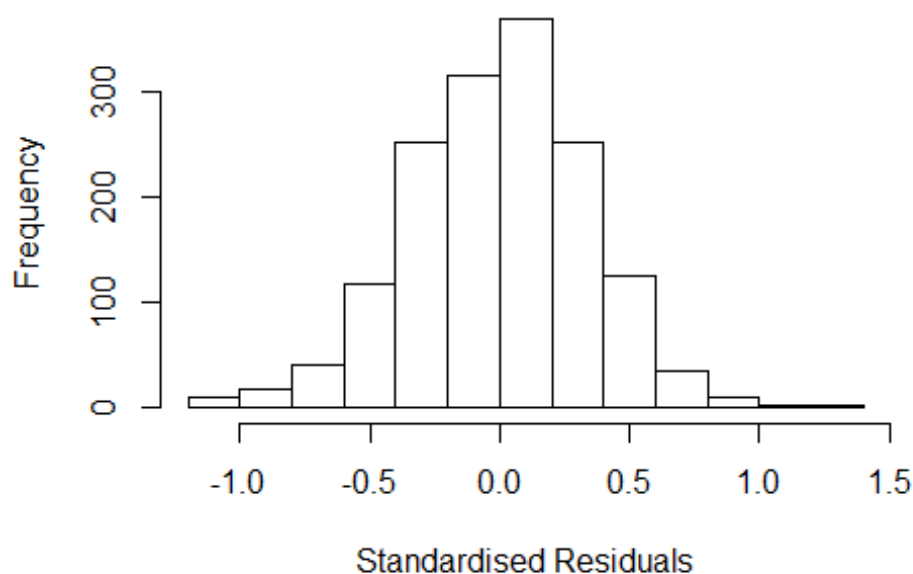
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20901 -0.22173  0.00752  0.21763  1.39775
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0584    0.0779   13.59 < 2e-16 ***
## FirstAuthorFemale1 -0.0255    0.0176   -1.45  0.14796
## LastAuthorFemale1 -0.0282    0.0223   -1.26  0.20649
## UniqueAuthors2     0.1716    0.0709    2.42  0.01555 *
## UniqueAuthors3     0.2694    0.0697    3.87  0.00012 ***
## UniqueAuthors4     0.2951    0.0698    4.23  2.5e-05 ***
## UniqueAuthors5     0.3383    0.0687    4.93  9.3e-07 ***
## Year1997          -0.0979    0.0522   -1.87  0.06100 .
## Year1998          -0.1483    0.0597   -2.48  0.01310 *
## Year1999          -0.1777    0.0534   -3.33  0.00090 ***
```

```

## Year2000          -0.1371      0.0648   -2.12  0.03435 *
## Year2001          -0.2172      0.0757   -2.87  0.00418 **
## Year2002          -0.1877      0.0490   -3.83  0.00013 ***
## Year2003          -0.1830      0.0524   -3.49  0.00049 ***
## Year2004          -0.2159      0.0508   -4.25  2.3e-05 ***
## Year2005          -0.1873      0.0499   -3.76  0.00018 ***
## Year2006          -0.1850      0.0552   -3.35  0.00083 ***
## Year2007          -0.1976      0.0501   -3.94  8.4e-05 ***
## Year2008          -0.2184      0.0523   -4.17  3.2e-05 ***
## Year2009          -0.1391      0.0499   -2.79  0.00540 **
## Year2010          -0.1794      0.0519   -3.46  0.00056 ***
## Year2011          -0.1706      0.0504   -3.39  0.00073 ***
## Year2012          -0.2331      0.0596   -3.91  9.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.327
## Multiple R-squared:  0.0716, Adjusted R-squared:  0.0582
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1421 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0276 0.8750 0.9510 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          6.47e-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 1.039 1 1.019
## LastAuthorFemale 1.081 1 1.040
## Year 1.122 16 1.004

```


Residuals from first and last author



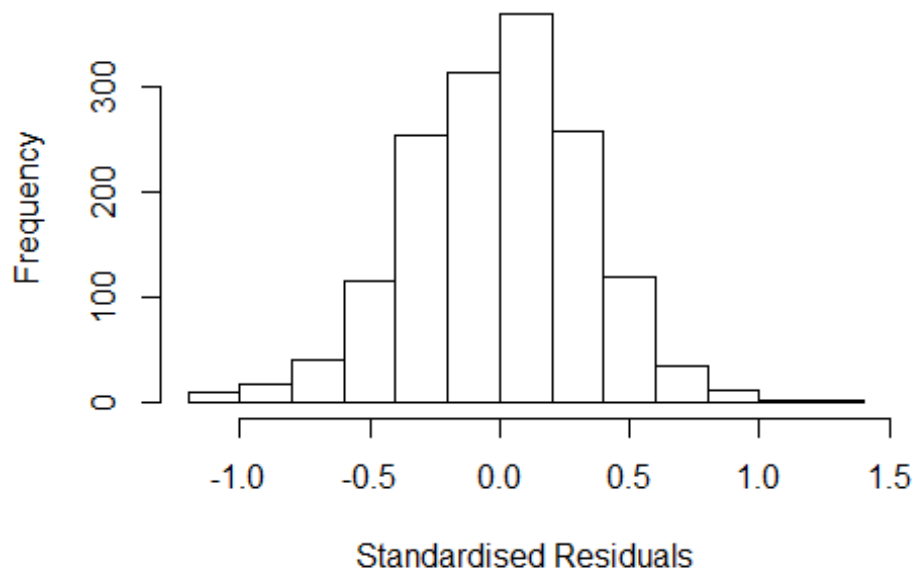
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1483 -0.2342  0.0139  0.2318  1.3224
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3092     0.0418   31.35 < 2e-16 ***
## FirstAuthorFemale1 -0.0167     0.0178   -0.94  0.34735
## LastAuthorFemale1  -0.0389     0.0232   -1.68  0.09341 .
## Year1997          -0.0860     0.0527   -1.63  0.10296
## Year1998          -0.1608     0.0628   -2.56  0.01048 *
## Year1999          -0.1651     0.0524   -3.15  0.00165 **
## Year2000          -0.1198     0.0672   -1.78  0.07484 .
## Year2001          -0.2312     0.0792   -2.92  0.00357 **
## Year2002          -0.1719     0.0510   -3.37  0.00076 ***
## Year2003          -0.1669     0.0540   -3.09  0.00204 **
## Year2004          -0.1959     0.0523   -3.75  0.00019 ***
## Year2005          -0.1668     0.0511   -3.27  0.00112 **
```

```

## Year2006          -0.1494      0.0549   -2.72  0.00656 **
## Year2007          -0.1678      0.0513   -3.27  0.00109 **
## Year2008          -0.1826      0.0530   -3.44  0.00059 ***
## Year2009          -0.1062      0.0516   -2.06  0.03984 *
## Year2010          -0.1611      0.0534   -3.02  0.00260 **
## Year2011          -0.1351      0.0517   -2.61  0.00908 **
## Year2012          -0.1878      0.0594   -3.16  0.00158 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0215, Adjusted R-squared:  0.00992
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1433 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0885 0.8770 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      6.47e-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 1.039 1      1.020
## Year      1.039 16      1.001

```

Residuals from first author



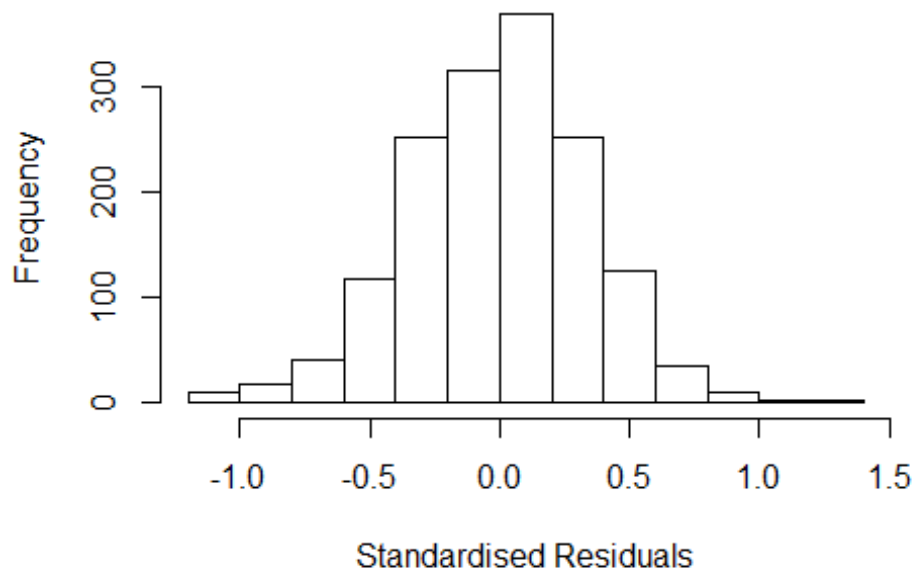
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1466 -0.2318 0.0147 0.2328 1.3253
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3032 0.0419 31.09 < 2e-16 ***
## FirstAuthorFemale1 -0.0179 0.0178 -1.00 0.31575
## Year1997 -0.0854 0.0529 -1.61 0.10664
## Year1998 -0.1566 0.0628 -2.49 0.01277 *
## Year1999 -0.1656 0.0526 -3.15 0.00167 **
## Year2000 -0.1170 0.0675 -1.73 0.08332 .
## Year2001 -0.2275 0.0791 -2.88 0.00409 **
## Year2002 -0.1730 0.0511 -3.38 0.00073 ***
## Year2003 -0.1691 0.0542 -3.12 0.00185 **
## Year2004 -0.1954 0.0522 -3.74 0.00019 ***
## Year2005 -0.1670 0.0513 -3.26 0.00115 **
## Year2006 -0.1534 0.0548 -2.80 0.00518 **
```

```

## Year2007          -0.1713      0.0512   -3.34  0.00085 ***
## Year2008          -0.1814      0.0532   -3.41  0.00066 ***
## Year2009          -0.1077      0.0519   -2.08  0.03797 *
## Year2010          -0.1627      0.0536   -3.04  0.00243 **
## Year2011          -0.1383      0.0518   -2.67  0.00764 **
## Year2012          -0.1904      0.0594   -3.21  0.00137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.00872
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1428 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0861 0.8770 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      6.47e-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 1.08 1      1.039
## Year      1.08 16      1.002

```

Residuals from last author



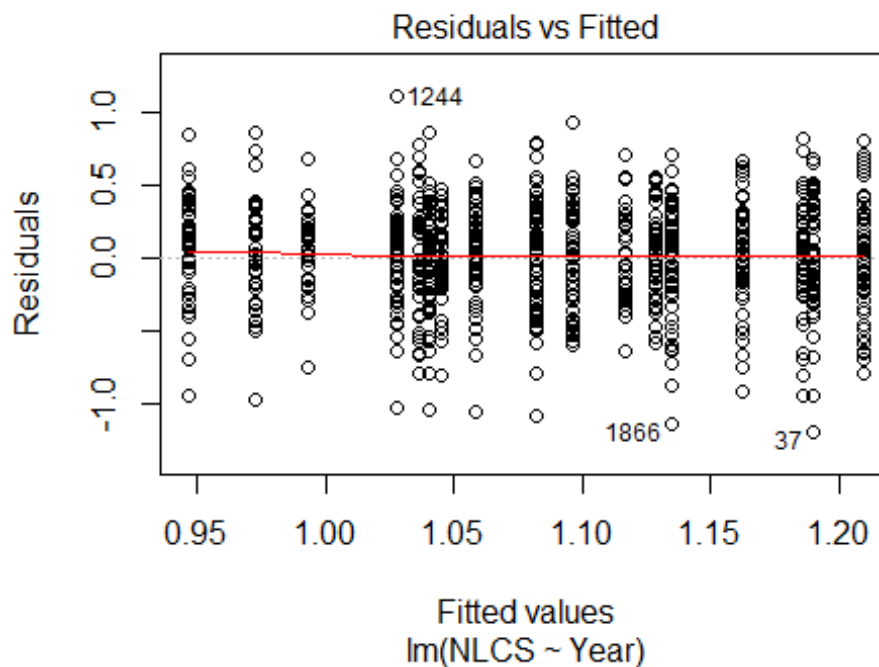
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1416 -0.2356 0.0132 0.2374 1.3124
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3010 0.0410 31.71 < 2e-16 ***
## LastAuthorFemale1 -0.0397 0.0231 -1.72 0.08620 .
## Year1997 -0.0839 0.0528 -1.59 0.11242
## Year1998 -0.1594 0.0631 -2.53 0.01165 *
## Year1999 -0.1627 0.0524 -3.10 0.00194 **
## Year2000 -0.1189 0.0675 -1.76 0.07827 .
## Year2001 -0.2269 0.0793 -2.86 0.00427 **
## Year2002 -0.1707 0.0511 -3.34 0.00086 ***
## Year2003 -0.1660 0.0544 -3.05 0.00234 **
## Year2004 -0.1959 0.0524 -3.74 0.00019 ***
## Year2005 -0.1666 0.0514 -3.24 0.00120 **
## Year2006 -0.1467 0.0549 -2.67 0.00763 **
```

```

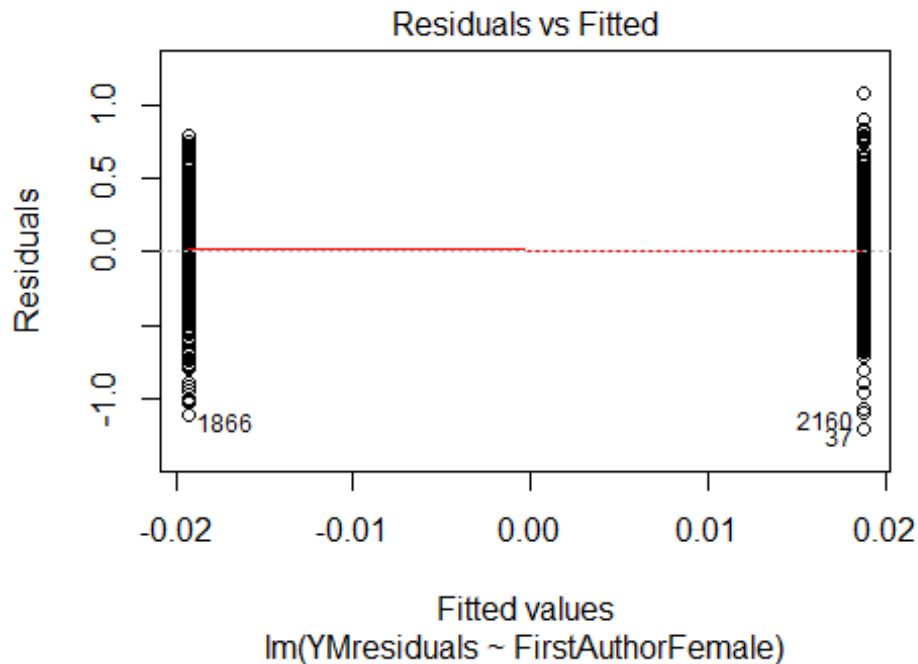
## Year2007          -0.1660      0.0513   -3.24  0.00124 **
## Year2008          -0.1819      0.0533   -3.41  0.00066 ***
## Year2009          -0.1063      0.0519   -2.05  0.04055 *
## Year2010          -0.1606      0.0536   -3.00  0.00278 **
## Year2011          -0.1345      0.0519   -2.59  0.00959 **
## Year2012          -0.1871      0.0593   -3.15  0.00164 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0209, Adjusted R-squared:  0.00999
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1423 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0963 0.8770 0.9490 0.9060 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      6.47e-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: 1545"
## [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
## 106 110 103 95 91 99 93 75 68 73 80 85 94 103 104
## 2011 2012
## 105 105
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 48 42 45 41 30 48 41 34 53 48 52 59 64 63
## 2011 2012

```

```
## 64 60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 41 32 39 39 25 44 38 30 48 42 46 52 55 57
## 2011 2012
## 56 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
```



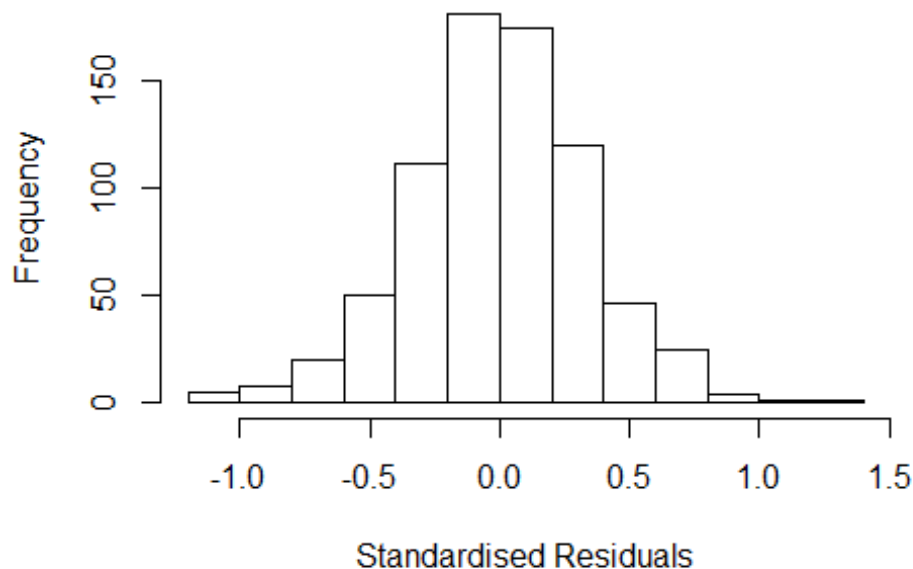
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.018, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 5.21906220885266"
## [1] "Male first author team size 2018 geometric mean: 4.02403215906174"
## 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 = 720, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.63953688611057"
## [1] "Male last author team size 2018 geometric mean: 4.5303082763943"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.184	1	1.088
LastAuthorFemale	1.146	1	1.071
UniqueAuthors	1.764	4	1.073
Year	1.922	16	1.021

Residuals from first and last author and team size



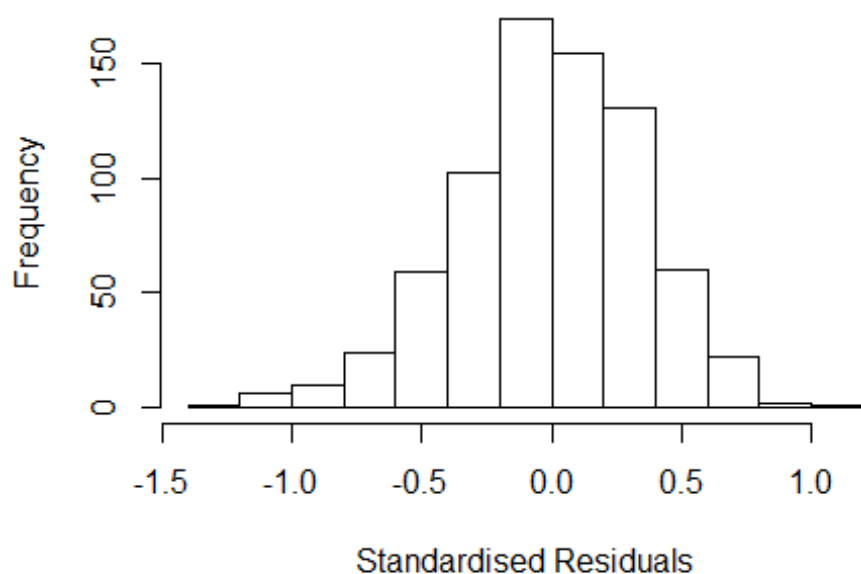
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1819 -0.2102 -0.0021  0.2168  1.2385
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9996    0.0832   12.01 < 2e-16 ***
## FirstAuthorFemale1 -0.0446    0.0257   -1.74  0.08313 .
## LastAuthorFemale1  0.0107    0.0305    0.35  0.72501
## UniqueAuthors2    0.1715    0.0694    2.47  0.01372 *
## UniqueAuthors3    0.2170    0.0683    3.18  0.00155 **
## UniqueAuthors4    0.2953    0.0696    4.25  2.5e-05 ***
## UniqueAuthors5    0.3986    0.0644    6.19  1.0e-09 ***
## Year1997         -0.0569    0.0700   -0.81  0.41654
## Year1998         -0.2298    0.0964   -2.38  0.01738 *
## Year1999         -0.2030    0.0673   -3.02  0.00263 **
```

```

## Year2000          -0.2615      0.0783   -3.34  0.00088 ***
## Year2001          -0.2246      0.0824   -2.73  0.00657 **
## Year2002          -0.1637      0.0682   -2.40  0.01668 *
## Year2003          -0.2900      0.0845   -3.43  0.00063 ***
## Year2004          -0.1596      0.0702   -2.27  0.02336 *
## Year2005          -0.1928      0.0742   -2.60  0.00953 **
## Year2006          -0.2191      0.0776   -2.82  0.00487 **
## Year2007          -0.1142      0.0694   -1.64  0.10044
## Year2008          -0.1946      0.0709   -2.75  0.00620 **
## Year2009          -0.0832      0.0722   -1.15  0.24941
## Year2010          -0.0875      0.0735   -1.19  0.23452
## Year2011          -0.1052      0.0745   -1.41  0.15834
## Year2012          -0.1583      0.0720   -2.20  0.02815 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.318
## Multiple R-squared:  0.161, Adjusted R-squared:  0.136
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 675 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0959 0.8710 0.9500 0.8970 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          1.34e-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 1.073 1 1.036
## LastAuthorFemale 1.143 1 1.069
## Year 1.186 16 1.005

```

Residuals from first and last author



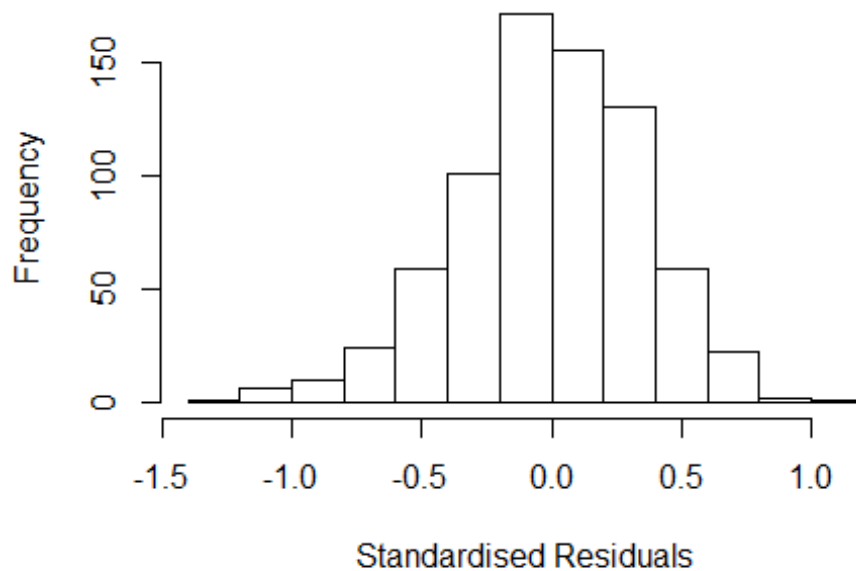
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20549 -0.22432 -0.00116 0.23007 1.07231
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22207 0.05040 24.25 <2e-16 ***
## FirstAuthorFemale1 -0.04645 0.02663 -1.74 0.0815 .
## LastAuthorFemale1 -0.01658 0.03361 -0.49 0.6219
## Year1997 -0.03345 0.07157 -0.47 0.6404
## Year1998 -0.23509 0.09405 -2.50 0.0126 *
## Year1999 -0.17939 0.06846 -2.62 0.0090 **
## Year2000 -0.22842 0.07443 -3.07 0.0022 **
## Year2001 -0.19866 0.07672 -2.59 0.0098 **
## Year2002 -0.06196 0.06662 -0.93 0.3526
## Year2003 -0.19946 0.08905 -2.24 0.0254 *
## Year2004 -0.08657 0.07482 -1.16 0.2476
## Year2005 -0.16338 0.07458 -2.19 0.0288 *
```

```

## Year2006      -0.11682    0.07635   -1.53    0.1265
## Year2007      -0.07642    0.06546   -1.17    0.2434
## Year2008      -0.11311    0.06667   -1.70    0.0902 .
## Year2009       0.00180    0.07000    0.03    0.9795
## Year2010      -0.03695    0.07201   -0.51    0.6081
## Year2011      -0.00694    0.07377   -0.09    0.9251
## Year2012      -0.11953    0.07146   -1.67    0.0948 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0512, Adjusted R-squared:  0.0277
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 676 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.186  0.870   0.949   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      1.34e-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 1.053 1      1.026
## Year              1.053 16      1.002

```

Residuals from first author



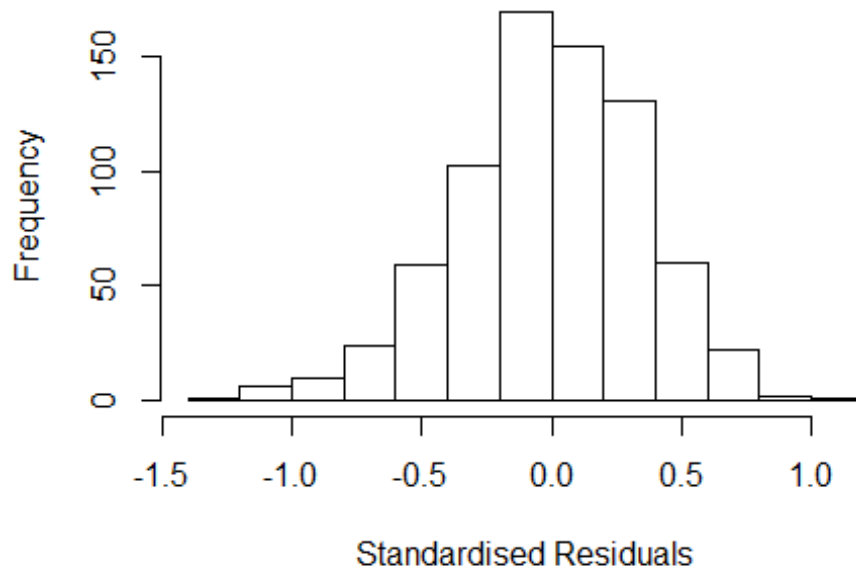
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22145 -0.22068 -0.00137 0.22739 1.07492
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22145 0.05049 24.19 <2e-16 ***
## FirstAuthorFemale1 -0.04892 0.02641 -1.85 0.0644 .
## Year1997 -0.03422 0.07174 -0.48 0.6335
## Year1998 -0.23770 0.09368 -2.54 0.0114 *
## Year1999 -0.18279 0.06816 -2.68 0.0075 **
## Year2000 -0.22825 0.07432 -3.07 0.0022 **
## Year2001 -0.19870 0.07666 -2.59 0.0097 **
## Year2002 -0.06223 0.06681 -0.93 0.3519
## Year2003 -0.20272 0.08975 -2.26 0.0242 *
## Year2004 -0.08968 0.07480 -1.20 0.2309
## Year2005 -0.16536 0.07473 -2.21 0.0272 *
## Year2006 -0.11790 0.07651 -1.54 0.1238
```

```

## Year2007      -0.07801    0.06580   -1.19    0.2362
## Year2008      -0.11715    0.06663   -1.76    0.0791 .
## Year2009      -0.00225    0.06904   -0.03    0.9740
## Year2010      -0.03803    0.07228   -0.53    0.5990
## Year2011      -0.00912    0.07385   -0.12    0.9018
## Year2012      -0.12255    0.07131   -1.72    0.0861 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0509, Adjusted R-squared:  0.0287
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 677 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.172  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      1.34e-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 1.121 1      1.059
## Year      1.121 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
## -1.17e+00 -2.28e-01  7.42e-05  2.36e-01  1.10e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19982    0.04956   24.21  <2e-16 ***
## LastAuthorFemale1 -0.02665    0.03329   -0.80   0.4237
## Year1997       -0.03149    0.07160   -0.44   0.6602
## Year1998       -0.22717    0.09491   -2.39   0.0169 *
## Year1999       -0.17144    0.06772   -2.53   0.0116 *
## Year2000       -0.22853    0.07560   -3.02   0.0026 **
## Year2001       -0.20365    0.07692   -2.65   0.0083 **
## Year2002       -0.05796    0.06730   -0.86   0.3894
## Year2003       -0.19741    0.08906   -2.22   0.0270 *
## Year2004       -0.08382    0.07532   -1.11   0.2661
## Year2005       -0.16440    0.07478   -2.20   0.0282 *
## Year2006       -0.11022    0.07658   -1.44   0.1505
```

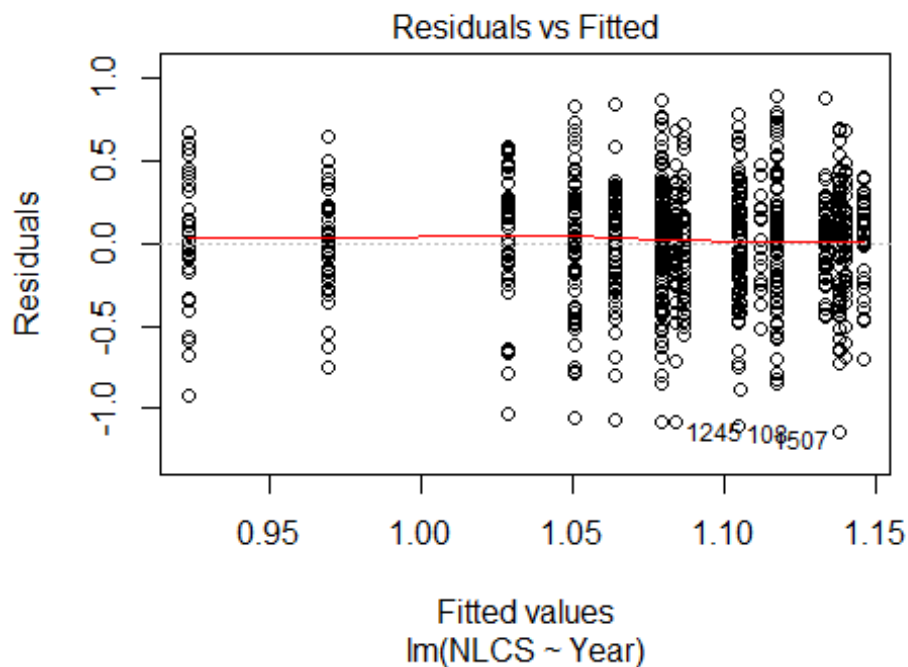
```

## Year2007          -0.07698      0.06521    -1.18    0.2382
## Year2008          -0.11328      0.06674    -1.70    0.0901 .
## Year2009           0.00163      0.06981     0.02    0.9813
## Year2010          -0.03386      0.07173    -0.47    0.6371
## Year2011          -0.00672      0.07343    -0.09    0.9271
## Year2012          -0.11601      0.07145    -1.62    0.1049
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.0471, Adjusted R-squared:  0.0248
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 681 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.212  0.871  0.949  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      1.34e-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: 744"
## [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
##   86   88   91   79   67   82   75   70   67   69   88   94   99   87   85
## 2011 2012
##  123   94
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   43   53   46   41   29   17   51   46   40   49   58   62   67   60   62
## 2011 2012

```



```
##      84      63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   40   49   39   37   24   13   49   42   38   43   56   54   64   55   57
## 2011 2012
##   77   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 = 47, df = 16, p-value = 8e-05
```



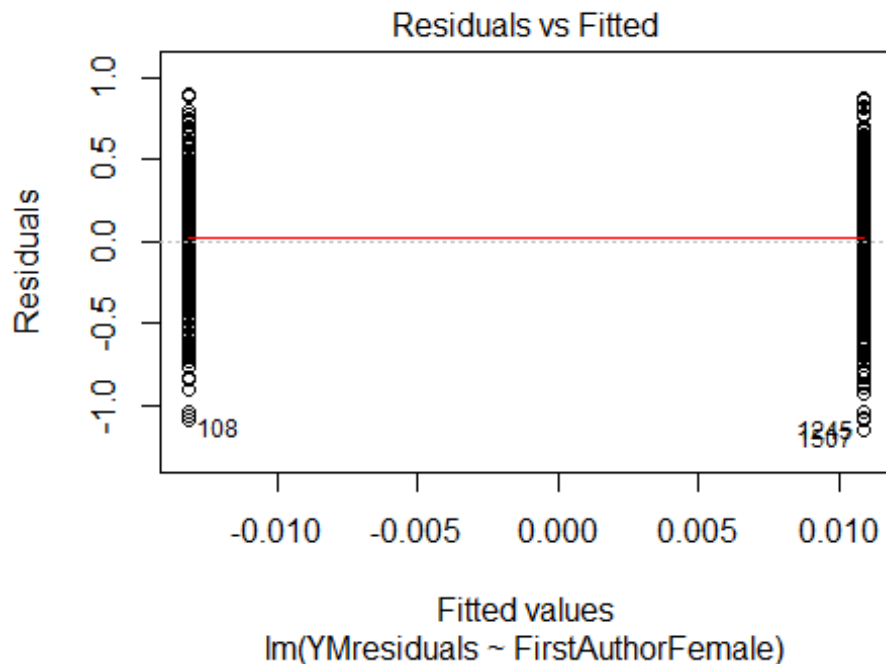
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0035, df = 1, p-value = 1

## [1] "Female first author team size 2018 geometric mean: 5.7588522979094"
## [1] "Male first author team size 2018 geometric mean: 5.32293132721702"

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

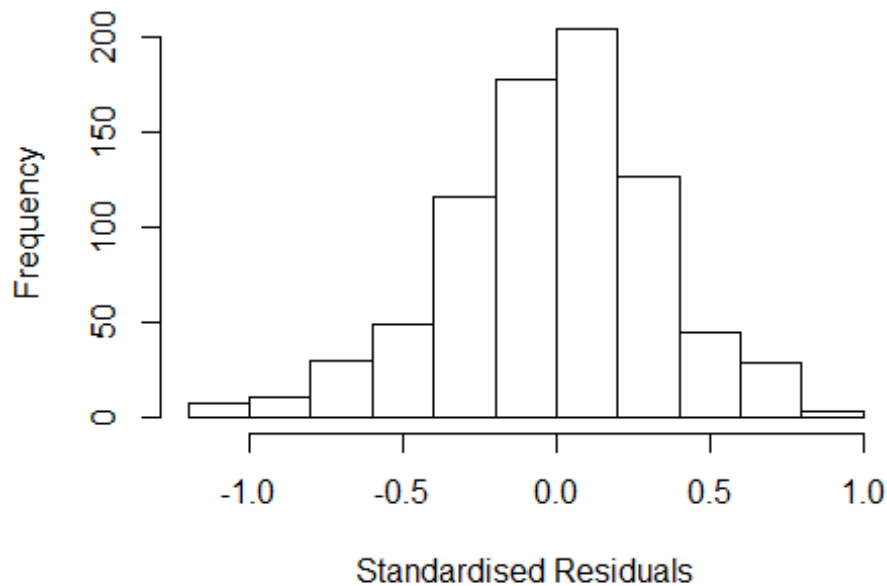
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 560, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.62265149623932"
## [1] "Male last author team size 2018 geometric mean: 5.60639311538178"

## 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 = 610, 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 1.133  1      1.065
## LastAuthorFemale  1.130  1      1.063
## UniqueAuthors    1.854  4      1.080
## Year              2.041 16      1.023
```

Residuals from first and last author and team size



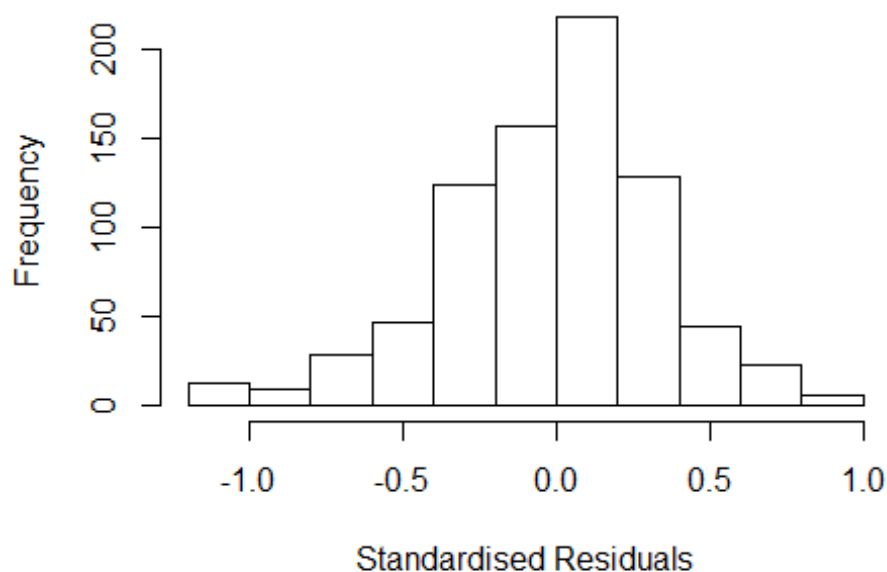
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1774 -0.2155 0.0115 0.2046 0.8932
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07763 0.12511 8.61 <2e-16 ***
## FirstAuthorFemale1 -0.03606 0.02585 -1.39 0.1635
## LastAuthorFemale1 -0.00531 0.03209 -0.17 0.8687
## UniqueAuthors2 0.06628 0.12348 0.54 0.5916
## UniqueAuthors3 0.03859 0.11762 0.33 0.7429
## UniqueAuthors4 0.11122 0.11983 0.93 0.3536
## UniqueAuthors5 0.13550 0.11627 1.17 0.2442
## Year1997 -0.04589 0.07894 -0.58 0.5612
## Year1998 -0.09145 0.07747 -1.18 0.2382
## Year1999 -0.18286 0.06910 -2.65 0.0083 **
```

```

## Year2000          0.00807    0.08274    0.10    0.9223
## Year2001         -0.03214    0.10350   -0.31    0.7563
## Year2002         -0.03204    0.05965   -0.54    0.5914
## Year2003         -0.06727    0.06438   -1.04    0.2964
## Year2004         -0.23100    0.08167   -2.83    0.0048 **
## Year2005         -0.08367    0.08141   -1.03    0.3044
## Year2006         -0.04487    0.07033   -0.64    0.5237
## Year2007         -0.06998    0.06098   -1.15    0.2515
## Year2008         -0.09551    0.06132   -1.56    0.1197
## Year2009         -0.01140    0.07570   -0.15    0.8803
## Year2010         -0.04911    0.06791   -0.72    0.4698
## Year2011         -0.05658    0.06586   -0.86    0.3906
## Year2012         -0.08563    0.07367   -1.16    0.2455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.317
## Multiple R-squared:  0.0452, Adjusted R-squared:  0.0179
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.137  0.867  0.947   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      1.26e-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 1.131 1      1.064
## LastAuthorFemale  1.074 1      1.036
## Year              1.202 16      1.006

```

Residuals from first and last author



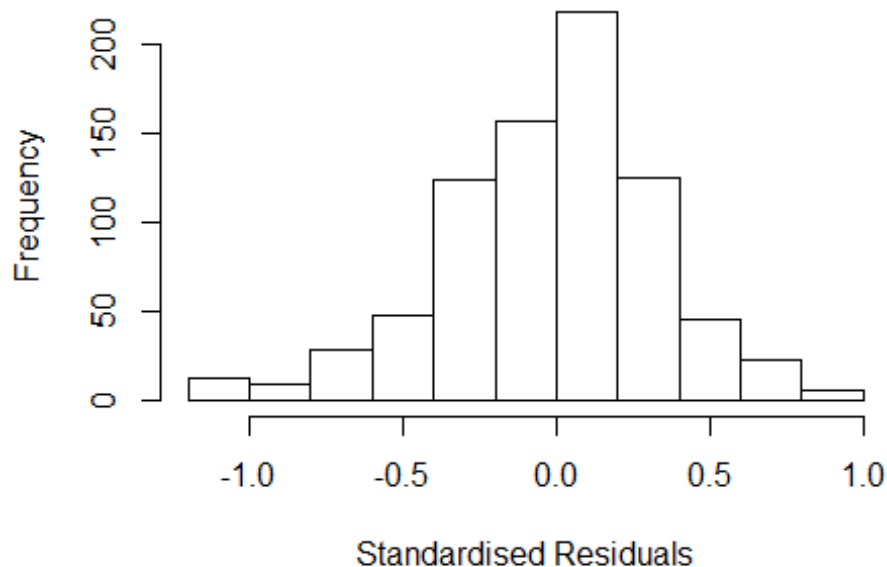
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1800 -0.2303  0.0224  0.2015  0.9002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.172151   0.046511   25.20  <2e-16 ***
## FirstAuthorFemale1 -0.033092   0.025798   -1.28   0.2000
## LastAuthorFemale1 -0.008151   0.031400   -0.26   0.7952
## Year1997         -0.047497   0.077651   -0.61   0.5409
## Year1998         -0.096787   0.072231   -1.34   0.1806
## Year1999         -0.186521   0.066899   -2.79   0.0054 **
## Year2000          0.000619   0.079725    0.01   0.9938
## Year2001         -0.024157   0.099907   -0.24   0.8090
## Year2002         -0.025222   0.057320   -0.44   0.6600
## Year2003         -0.056742   0.061827   -0.92   0.3590
## Year2004         -0.224058   0.077368   -2.90   0.0039 **
## Year2005         -0.068662   0.076189   -0.90   0.3678
```

```

## Year2006      -0.036019    0.068152   -0.53    0.5973
## Year2007      -0.058689    0.057699   -1.02    0.3094
## Year2008      -0.074269    0.057521   -1.29    0.1970
## Year2009       0.007868    0.074812    0.11    0.9163
## Year2010      -0.035465    0.064527   -0.55    0.5827
## Year2011      -0.047383    0.063847   -0.74    0.4582
## Year2012      -0.078614    0.070669   -1.11    0.2663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.00698
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.141  0.868  0.952  0.891  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      1.26e-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 1.127 1      1.061
## Year              1.127 16      1.004

```

Residuals from first author



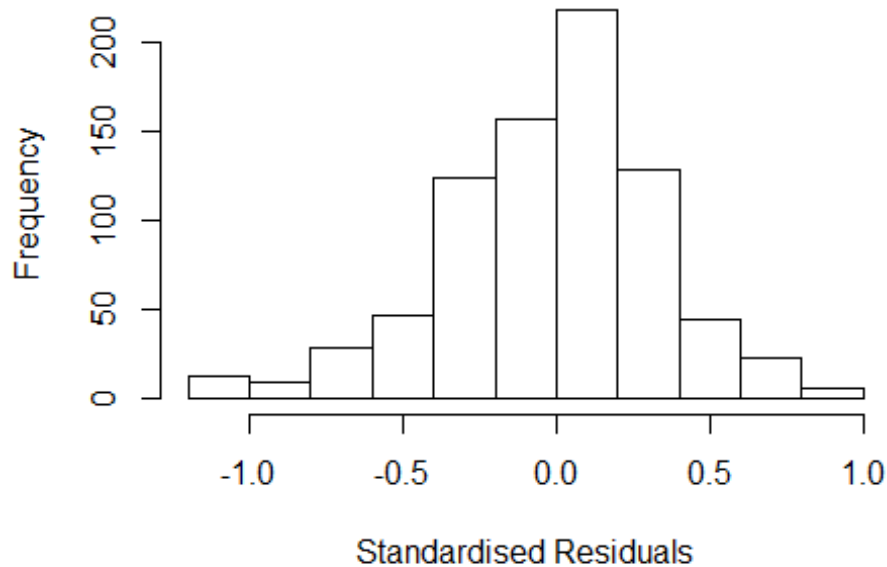
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1788 -0.2308 0.0223 0.1998 0.9016
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17114 0.04630 25.30 <2e-16 ***
## FirstAuthorFemale1 -0.03377 0.02575 -1.31 0.1901
## Year1997 -0.04761 0.07767 -0.61 0.5401
## Year1998 -0.09680 0.07229 -1.34 0.1809
## Year1999 -0.18688 0.06687 -2.79 0.0053 **
## Year2000 0.00153 0.07968 0.02 0.9846
## Year2001 -0.02385 0.10042 -0.24 0.8123
## Year2002 -0.02497 0.05737 -0.44 0.6635
## Year2003 -0.05731 0.06180 -0.93 0.3540
## Year2004 -0.22434 0.07730 -2.90 0.0038 **
## Year2005 -0.06857 0.07609 -0.90 0.3678
## Year2006 -0.03525 0.06822 -0.52 0.6055
```

```

## Year2007          -0.05916    0.05764   -1.03    0.3050
## Year2008          -0.07360    0.05744   -1.28    0.2005
## Year2009           0.00765    0.07457    0.10    0.9183
## Year2010          -0.03589    0.06447   -0.56    0.5779
## Year2011          -0.04750    0.06393   -0.74    0.4577
## Year2012          -0.07947    0.07074   -1.12    0.2616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.318
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.00825
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 726 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.141  0.868  0.952  0.891  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      1.26e-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 1.069 1          1.034
## Year              1.069 16          1.002

```


Residuals from last author



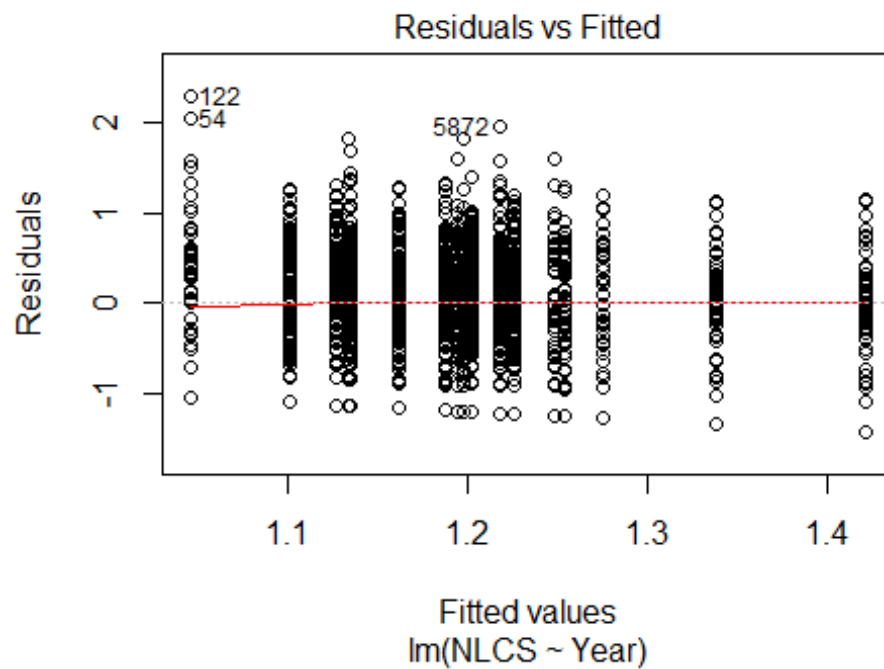
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1661 -0.2253 0.0193 0.1951 0.8933
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16100 0.04603 25.22 <2e-16 ***
## LastAuthorFemale1 -0.01258 0.03138 -0.40 0.6887
## Year1997 -0.04188 0.07787 -0.54 0.5909
## Year1998 -0.09972 0.07270 -1.37 0.1706
## Year1999 -0.18380 0.06775 -2.71 0.0068 **
## Year2000 0.00146 0.07883 0.02 0.9853
## Year2001 -0.02380 0.09803 -0.24 0.8083
## Year2002 -0.02726 0.05746 -0.47 0.6353
## Year2003 -0.05983 0.06197 -0.97 0.3346
## Year2004 -0.23297 0.07648 -3.05 0.0024 **
## Year2005 -0.06853 0.07648 -0.90 0.3705
## Year2006 -0.04494 0.06762 -0.66 0.5065
```

```

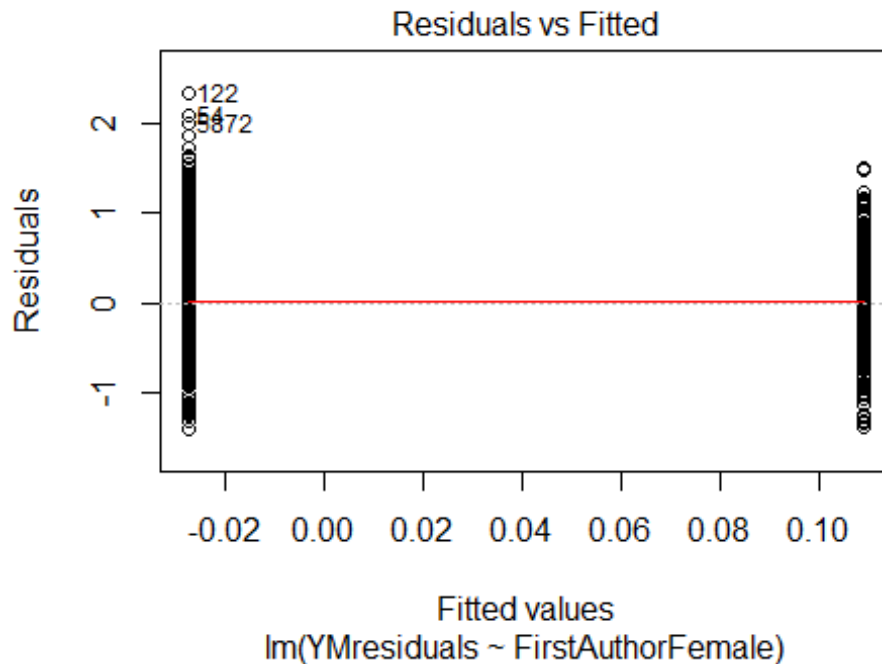
## Year2007          -0.06424      0.05725    -1.12    0.2622
## Year2008          -0.07791      0.05753    -1.35    0.1761
## Year2009           0.00514      0.07466     0.07    0.9451
## Year2010          -0.04086      0.06478    -0.63    0.5283
## Year2011          -0.04930      0.06363    -0.77    0.4387
## Year2012          -0.07914      0.07077    -1.12    0.2638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0272, Adjusted R-squared:  0.0059
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 735 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.154  0.866  0.954  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      1.26e-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: 795"
## [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
## 298 253 290 288 288 340 330 327 348 303 323 358 467 479 511
## 2011 2012
## 528 561
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 48 45 61 59 53 93 107 125 129 140 152 193 251 230
## 2011 2012

```

```
## 232 256
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 38 37 47 45 44 80 90 115 108 120 127 156 203 193
## 2011 2012
## 182 204
## [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 = 7e-06
```

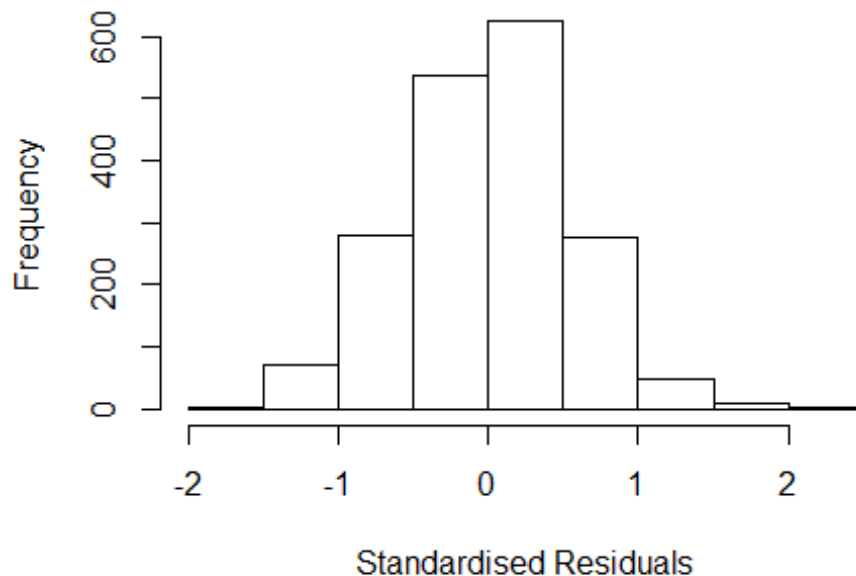


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.7, df = 1, p-value = 0.01
```



```
## [1] "Female first author team size 2018 geometric mean: 4.19441548717839"
## [1] "Male first author team size 2018 geometric mean: 3.56248152310157"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7400, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.16174002486725"
## [1] "Male last author team size 2018 geometric mean: 3.57754811286267"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6900, 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 1.064 1          1.031
## LastAuthorFemale  1.047 1          1.023
## UniqueAuthors     1.278 4          1.031
## Year               1.328 16         1.009
```

Residuals from first and last author and team size



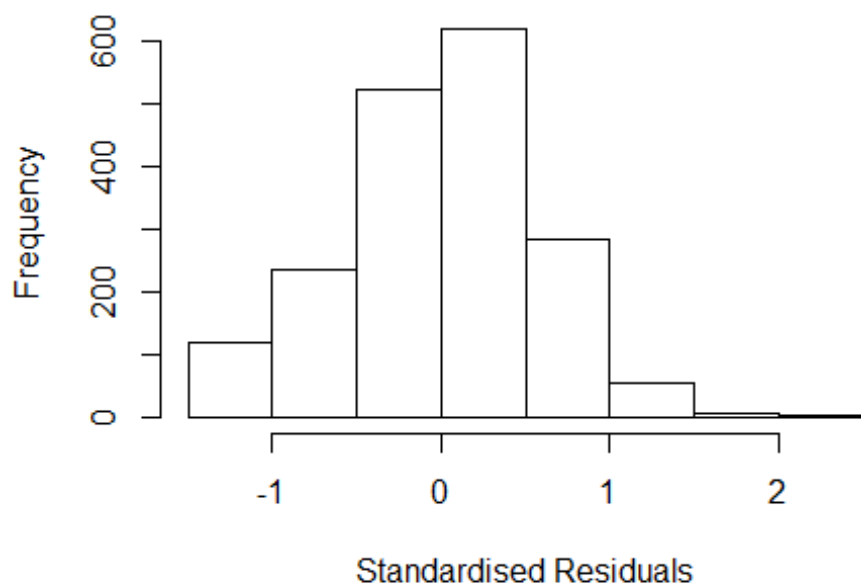
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6359 -0.3806 0.0293 0.3809 2.4600
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8840 0.1141 7.75 1.6e-14 ***
## FirstAuthorFemale1 0.1172 0.0316 3.71 0.00022 ***
## LastAuthorFemale1 0.0228 0.0360 0.63 0.52738
## UniqueAuthors2 0.2516 0.0510 4.93 8.8e-07 ***
## UniqueAuthors3 0.0972 0.0492 1.98 0.04835 *
## UniqueAuthors4 0.2227 0.0525 4.24 2.3e-05 ***
## UniqueAuthors5 0.3988 0.0499 8.00 2.3e-15 ***
## Year1997 0.1278 0.1628 0.79 0.43235
## Year1998 0.2359 0.1585 1.49 0.13675
## Year1999 0.2429 0.1465 1.66 0.09757 .
```

```

## Year2000          0.2761      0.1404      1.97  0.04934 *
## Year2001          0.3857      0.1352      2.85  0.00437 **
## Year2002          0.0150      0.1298      0.12  0.90808
## Year2003          0.1354      0.1287      1.05  0.29302
## Year2004          0.0227      0.1251      0.18  0.85590
## Year2005          0.0502      0.1220      0.41  0.68095
## Year2006          0.0102      0.1230      0.08  0.93422
## Year2007         -0.0227      0.1228     -0.18  0.85351
## Year2008          0.0445      0.1187      0.38  0.70763
## Year2009          0.1164      0.1143      1.02  0.30861
## Year2010          0.0934      0.1145      0.82  0.41482
## Year2011          0.0953      0.1157      0.82  0.41052
## Year2012          0.0595      0.1155      0.52  0.60636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0781, Adjusted R-squared:  0.067
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1703 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.013  0.871  0.950  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      5.41e-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 1.049 1      1.024
## LastAuthorFemale  1.039 1      1.019
## Year              1.069 16      1.002

```

Residuals from first and last author



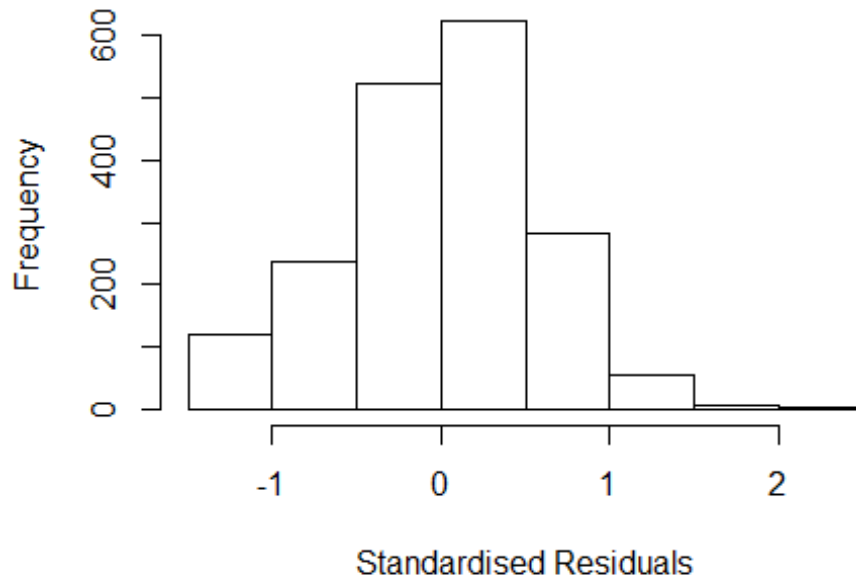
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4789 -0.3813 0.0281 0.3945 2.3246
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0194 0.1195 8.53 < 2e-16 ***
## FirstAuthorFemale1 0.1461 0.0321 4.55 5.6e-06 ***
## LastAuthorFemale1 0.0343 0.0369 0.93 0.3531
## Year1997 0.1604 0.1684 0.95 0.3410
## Year1998 0.2766 0.1656 1.67 0.0951 .
## Year1999 0.2554 0.1533 1.67 0.0959 .
## Year2000 0.2852 0.1487 1.92 0.0552 .
## Year2001 0.4252 0.1435 2.96 0.0031 **
## Year2002 0.0746 0.1395 0.53 0.5928
## Year2003 0.1497 0.1363 1.10 0.2723
## Year2004 0.0647 0.1346 0.48 0.6306
## Year2005 0.0713 0.1307 0.55 0.5852
```

```

## Year2006          0.0423      0.1326      0.32      0.7500
## Year2007          0.0412      0.1331      0.31      0.7568
## Year2008          0.1133      0.1286      0.88      0.3782
## Year2009          0.1781      0.1247      1.43      0.1535
## Year2010          0.1718      0.1246      1.38      0.1681
## Year2011          0.1747      0.1258      1.39      0.1651
## Year2012          0.1686      0.1252      1.35      0.1783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.0209
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1692 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0546 0.8660 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      5.41e-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 1.039 1      1.020
## 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
## -1.4512 -0.3815 0.0305 0.3923 2.3233
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0207 0.1197 8.52 < 2e-16 ***
## FirstAuthorFemale1 0.1489 0.0320 4.66 3.4e-06 ***
## Year1997 0.1589 0.1686 0.94 0.346
## Year1998 0.2816 0.1659 1.70 0.090 .
## Year1999 0.2600 0.1535 1.69 0.090 .
## Year2000 0.2850 0.1489 1.91 0.056 .
## Year2001 0.4271 0.1437 2.97 0.003 **
## Year2002 0.0793 0.1397 0.57 0.570
## Year2003 0.1513 0.1366 1.11 0.268
## Year2004 0.0676 0.1348 0.50 0.616
## Year2005 0.0747 0.1309 0.57 0.568
## Year2006 0.0448 0.1329 0.34 0.736
```

```

## Year2007          0.0437      0.1335      0.33      0.744
## Year2008          0.1165      0.1288      0.90      0.366
## Year2009          0.1828      0.1250      1.46      0.144
## Year2010          0.1770      0.1249      1.42      0.157
## Year2011          0.1791      0.1260      1.42      0.155
## Year2012          0.1724      0.1255      1.37      0.170
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.566
## Multiple R-squared:  0.03,   Adjusted R-squared:  0.021
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 165 weights are ~= 1. The remaining 1684 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0545 0.8640 0.9480 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      5.41e-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 1.035 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
## -1.5171 -0.3884  0.0344  0.3904  2.3023

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0417    0.1218   8.55 <2e-16 ***
## LastAuthorFemale1 0.0499    0.0372   1.34  0.1795
## Year1997          0.1503    0.1711   0.88  0.3799
## Year1998          0.2605    0.1692   1.54  0.1240
## Year1999          0.2542    0.1531   1.66  0.0970 .
## Year2000          0.2951    0.1490   1.98  0.0477 *
## Year2001          0.4256    0.1460   2.91  0.0036 **
## Year2002          0.0783    0.1411   0.55  0.5793
## Year2003          0.1586    0.1388   1.14  0.2534
## Year2004          0.0625    0.1370   0.46  0.6483
## Year2005          0.0760    0.1335   0.57  0.5691
## Year2006          0.0418    0.1350   0.31  0.7570
## Year2007          0.0543    0.1360   0.40  0.6897
## Year2008          0.1187    0.1314   0.90  0.3665
## Year2009          0.1879    0.1273   1.48  0.1401
## Year2010          0.1773    0.1273   1.39  0.1636
## Year2011          0.1909    0.1283   1.49  0.1371
## Year2012          0.1838    0.1276   1.44  0.1501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.0203, Adjusted R-squared:  0.0112
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1689 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0639 0.8640 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           5.41e-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: 1849"
## [1] ""

```

```

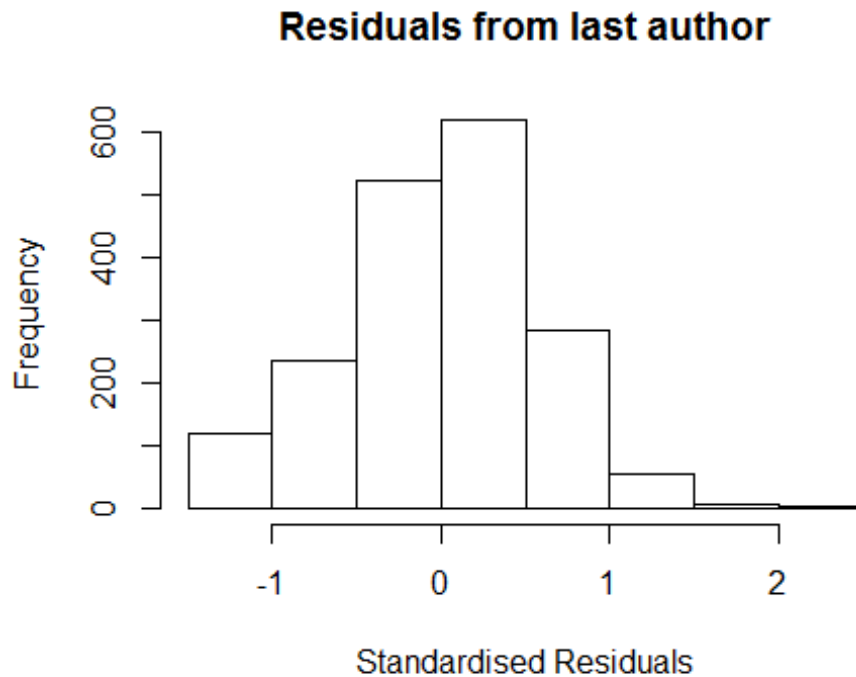
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2501"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 17 32 15 22 19 14 9 6 13 21 12 18 6 4
## 2011 2012
## 8 19
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 7 12 7 9 3 9 3 2 5 10 8 11 5 1
## 2011 2012
## 6 9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 6 10 7 9 3 9 2 2 3 9 7 9 5 1
## 2011 2012
## 2 5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.09273006856782"
## [1] "Male first author team size 2018 geometric mean: 1.5157165665104"

## 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 = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.83841628725254"
## [1] "Male last author team size 2018 geometric mean: 1.78179743628068"

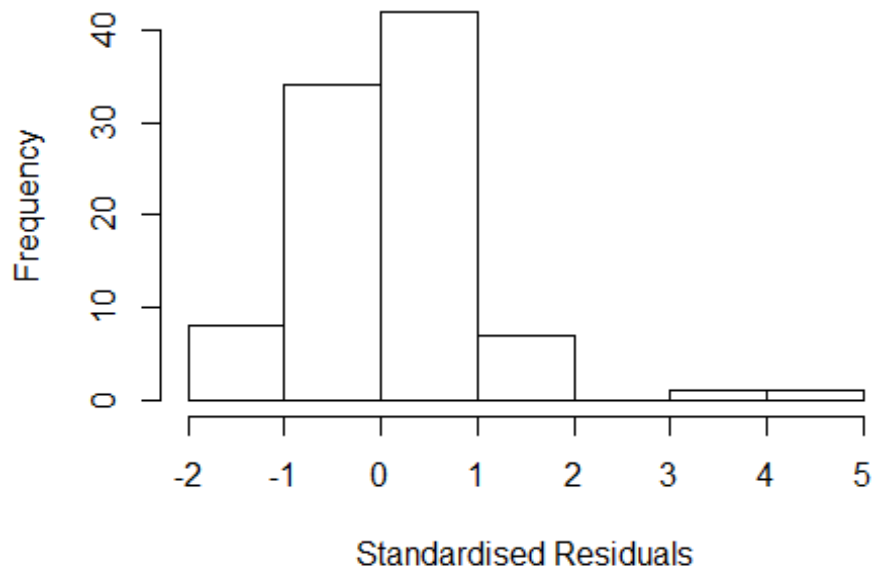
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 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 14.50  1      3.808
## LastAuthorFemale  21.52  1      4.639
## UniqueAuthors    115.15  4      1.810
## Year              960.00 16      1.239
```

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
## 27  4043158599 4.664 1997    2501      1    3.609
## 78 17444384494 5.611 1999    2501      1    4.541
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2439 -0.4166  0.0271  0.3651  4.5408
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.42916    0.23377     1.84  0.0706 .
## FirstAuthorFemale1 0.01130    0.22485     0.05  0.9600
## LastAuthorFemale1 0.17610    0.25702     0.69  0.4955
## UniqueAuthors2    0.69787    0.24167     2.89  0.0052 **
## UniqueAuthors3    0.69609    0.24774     2.81  0.0064 **
## UniqueAuthors4    0.63863    0.24015     2.66  0.0097 **
## UniqueAuthors5    0.67917    0.26408     2.57  0.0122 *
## Year1997        -0.01251    0.44976    -0.03  0.9779
## Year1998         0.55641    0.50500     1.10  0.2743
```

```

## Year1999      0.64103      0.47792      1.34      0.1842
## Year2000      0.13561      0.21823      0.62      0.5363
## Year2001      0.56225      0.59568      0.94      0.3485
## Year2002     -0.06906      0.36175     -0.19      0.8492
## Year2003      0.85936      0.99868      0.86      0.3925
## Year2004      0.98476      0.60652      1.62      0.1089
## Year2005      0.17764      0.33135      0.54      0.5936
## Year2006      0.55905      0.26413      2.12      0.0378 *
## Year2007     -0.14217      0.18506     -0.77      0.4449
## Year2008     -0.00855      0.18881     -0.05      0.9640
## Year2009      0.01579      0.36164      0.04      0.9653
## Year2010      0.33056      0.33066      1.00      0.3209
## Year2011      0.67600      0.31855      2.12      0.0374 *
## Year2012     -0.03480      0.30552     -0.11      0.9096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.709
## Multiple R-squared:  0.275, Adjusted R-squared:  0.0476
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(7,22) are outliers with |weight| = 0 ( < 0.0011);
## 13 weights are ~= 1. The remaining 78 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.581  0.851  0.961  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           1.08e-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 12.11 1           3.480
## LastAuthorFemale  19.06 1           4.366
## Year              57.38 16           1.135

```

```

## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 27  4043158599 4.664 1997      2501      1      3.806
## 78 17444384494 5.611 1999      2501      1      4.190
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.4210 -0.3887  0.0243  0.4533  4.1900
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.98309    0.32399   3.03  0.0033 **
## FirstAuthorFemale1 0.04573    0.24815   0.18  0.8543
## LastAuthorFemale1 0.11389    0.27718   0.41  0.6824
## Year1997        -0.12497    0.49482  -0.25  0.8013
## Year1998         0.36000    0.64966   0.55  0.5812
## Year1999         0.43791    0.81416   0.54  0.5923
## Year2000         0.17257    0.34772   0.50  0.6212
## Year2001         0.13766    0.93462   0.15  0.8833
## Year2002        -0.14570    0.36460  -0.40  0.6906
## Year2003         1.00241    0.98576   1.02  0.3125
## Year2004         0.77041    0.40691   1.89  0.0622 .
## Year2005         0.28089    0.38282   0.73  0.4654
## Year2006         0.45052    0.41909   1.08  0.2859
## Year2007        -0.20191    0.33487  -0.60  0.5484
## Year2008        -0.02742    0.32337  -0.08  0.9327
## Year2009        -0.00815    0.43751  -0.02  0.9852
## Year2010         0.50229    0.39438   1.27  0.2068
## Year2011         0.49879    0.41065   1.21  0.2284
## Year2012         0.08721    0.35872   0.24  0.8086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.755
## Multiple R-squared:  0.135, Adjusted R-squared:  -0.0748
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 2 observations c(7,22) are outliers with |weight| = 0 ( < 0.0011);
## 6 weights are ~1. The remaining 85 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.480  0.891  0.961   0.918   0.993   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00        5.00e-01      4.69e+00      1.00e-07

```

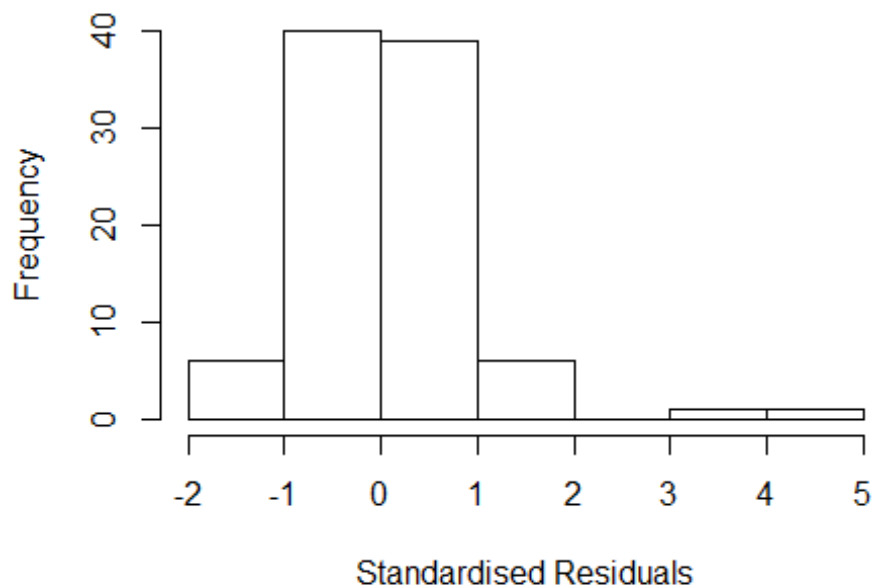


```
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          1.08e-03          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##          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): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps

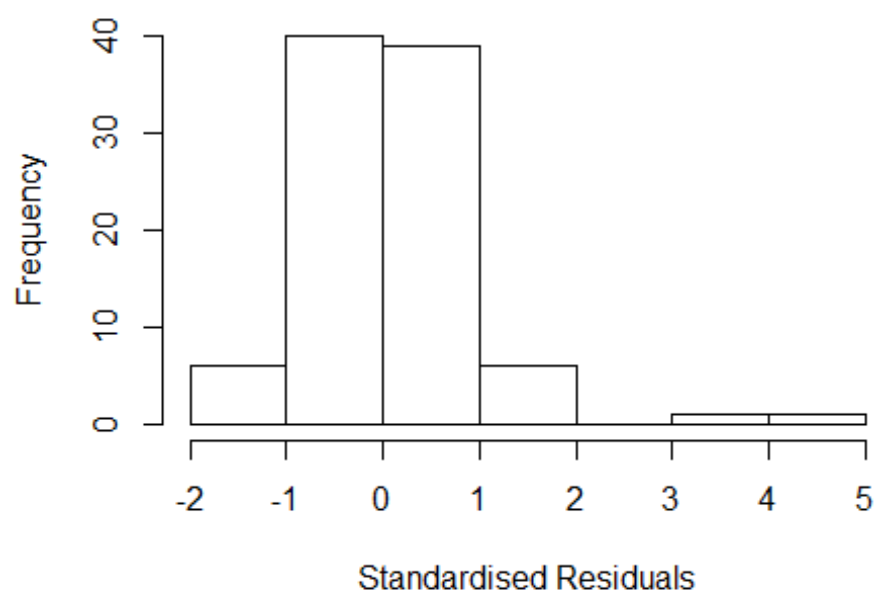
## Warning in lmrob.S(x, y, control = control, mf = mf): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps
```

Residuals from first and last author



```
## [1] "Regression 4: Last author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale    20  1          4.473
## Year                20 16          1.098
```

Residuals from last author



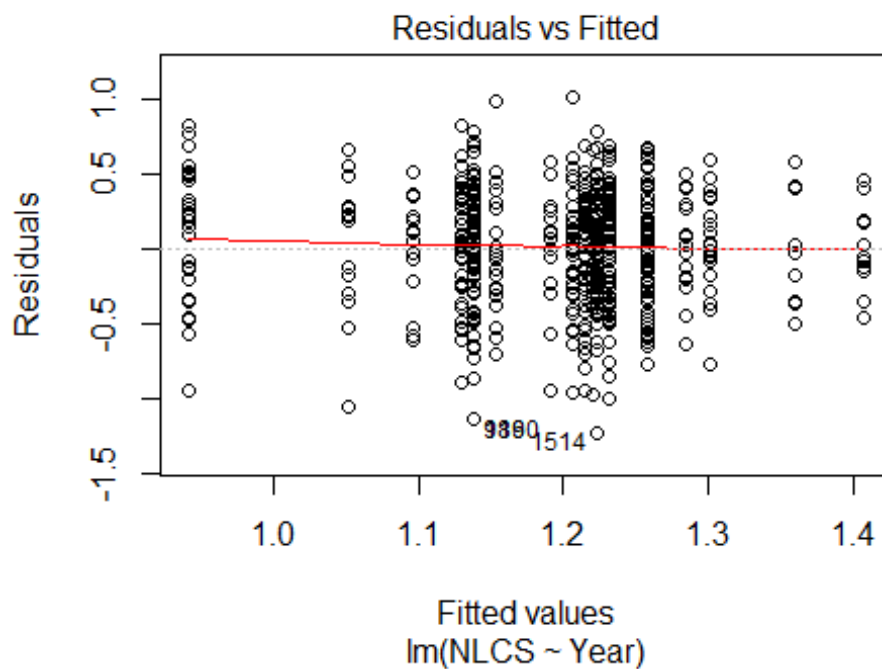
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 27  4043158599 4.664 1997    2501      1    3.806
## 78 17444384494 5.611 1999    2501      1    4.190
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4210 -0.4020  0.0192  0.4734  4.1900
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.99348    0.30981   3.21  0.002 **
## LastAuthorFemale1 0.12273    0.28360   0.43  0.666
## Year1997      -0.13546    0.48748  -0.28  0.782
## Year1998       0.34997    0.61624   0.57  0.572
## Year1999       0.42756    0.81207   0.53  0.600
## Year2000       0.16731    0.34580   0.48  0.630
## Year2001       0.12817    0.92741   0.14  0.890
## Year2002      -0.14692    0.36456  -0.40  0.688
## Year2003       0.99202    0.97354   1.02  0.311
## Year2004       0.76002    0.39568   1.92  0.059 .
## Year2005       0.28655    0.39052   0.73  0.465
```

```

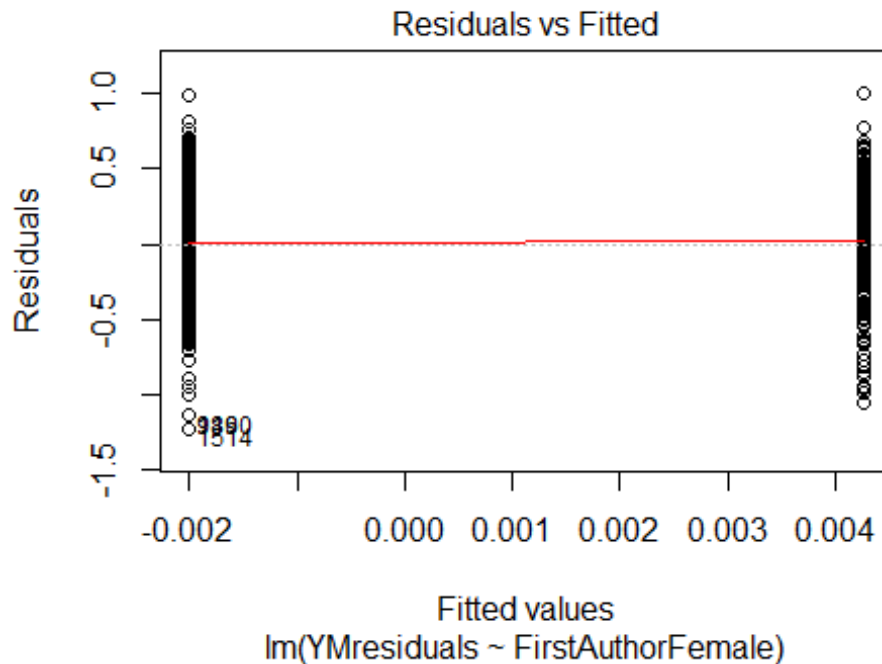
## Year2006      0.43976      0.39947      1.10      0.274
## Year2007     -0.19433      0.34281     -0.57      0.572
## Year2008     -0.02452      0.32538     -0.08      0.940
## Year2009     -0.00916      0.43998     -0.02      0.983
## Year2010      0.52879      0.35806      1.48      0.144
## Year2011      0.52529      0.37590      1.40      0.166
## Year2012      0.10315      0.35481      0.29      0.772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.757
## Multiple R-squared:  0.135, Adjusted R-squared:  -0.0615
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 2 observations c(7,22) are outliers with |weight| = 0 ( < 0.0011);
## 6 weights are ~ = 1. The remaining 85 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.482  0.889   0.964   0.918   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      1.08e-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: 93"
## [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
##   33   22   29   48   47   36   36   28   48   58   63   74   97  134  160
## 2011 2012
##   117  143
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 16 11 9 23 23 15 14 16 22 29 39 38 54 74 87
## 2011 2012
## 67 74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 8 6 20 21 10 13 15 20 27 37 31 45 63 75
## 2011 2012
## 62 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 = 14, df = 16, p-value = 0.6
```

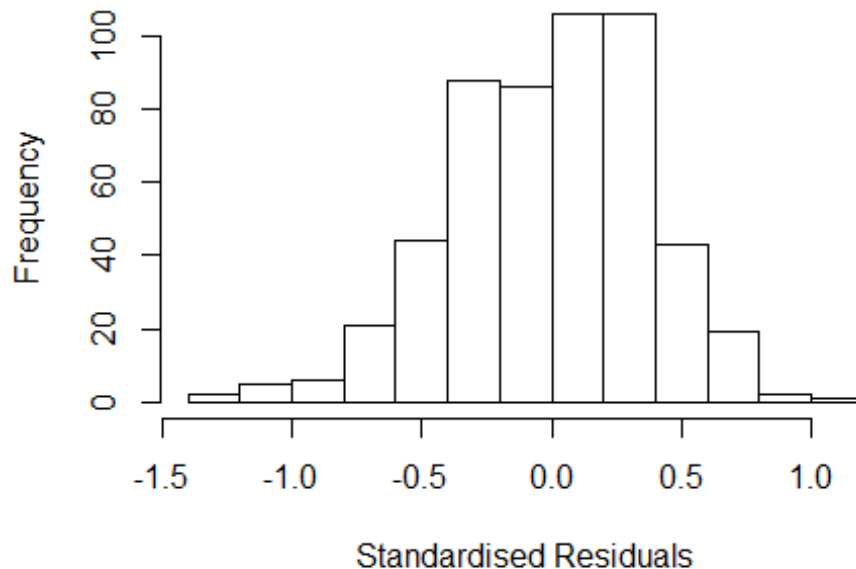


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.1, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 5.16140225519446"
## [1] "Male first author team size 2018 geometric mean: 4.41518440252618"
##
## 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: 4.16727167712065"
## [1] "Male last author team size 2018 geometric mean: 4.67405192309361"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 380, 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 1.289 1          1.135
## LastAuthorFemale  1.155 1          1.075
## UniqueAuthors    2.224 4          1.105
## Year              2.937 16          1.034
```

Residuals from first and last author and team size



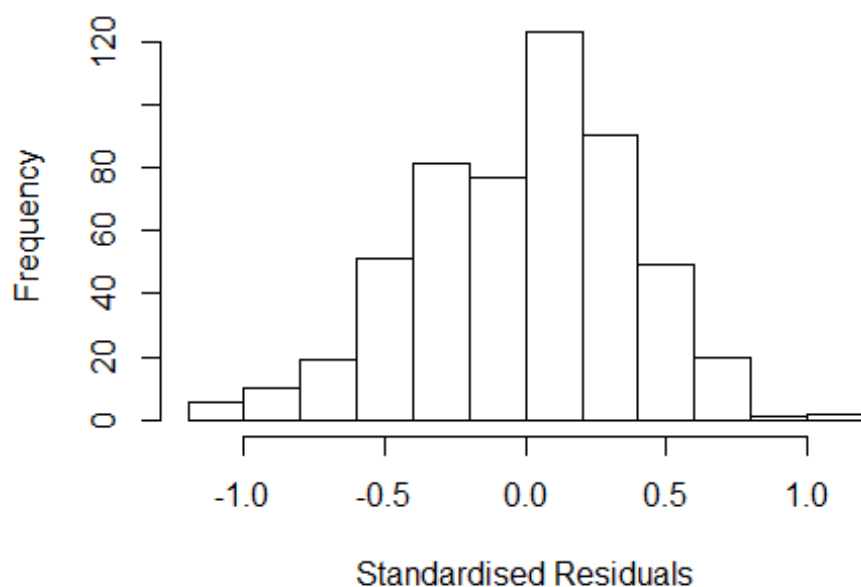
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2178 -0.2648 0.0183 0.2437 1.0843
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08586 0.11818 9.19 < 2e-16 ***
## FirstAuthorFemale1 -0.00878 0.04028 -0.22 0.82761
## LastAuthorFemale1 0.02735 0.04177 0.65 0.51290
## UniqueAuthors2 0.15957 0.09783 1.63 0.10349
## UniqueAuthors3 0.19690 0.09419 2.09 0.03707 *
## UniqueAuthors4 0.32815 0.09384 3.50 0.00051 ***
## UniqueAuthors5 0.32295 0.09095 3.55 0.00042 ***
## Year1997 0.21145 0.13460 1.57 0.11683
## Year1998 0.04971 0.18712 0.27 0.79061
## Year1999 -0.19469 0.13548 -1.44 0.15131
```

```

## Year2000      -0.05680      0.11993      -0.47      0.63600
## Year2001      -0.30893      0.18614      -1.66      0.09759 .
## Year2002      -0.04947      0.15579      -0.32      0.75097
## Year2003      -0.19275      0.13014      -1.48      0.13922
## Year2004      -0.06393      0.11201      -0.57      0.56842
## Year2005      -0.31930      0.15491      -2.06      0.03980 *
## Year2006      -0.19111      0.11224      -1.70      0.08924 .
## Year2007      -0.19309      0.11643      -1.66      0.09785 .
## Year2008      -0.11512      0.10769      -1.07      0.28559
## Year2009      -0.10465      0.10595      -0.99      0.32377
## Year2010      -0.19100      0.10389      -1.84      0.06658 .
## Year2011      -0.12562      0.10768      -1.17      0.24393
## Year2012      -0.16848      0.10450      -1.61      0.10752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.0889, Adjusted R-squared:  0.0493
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 486 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.288  0.886  0.949   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          1.89e-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 1.210 1          1.100
## LastAuthorFemale  1.107 1          1.052
## Year              1.315 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
## -1.1929 -0.2706 0.0281 0.2663 1.0354
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27264 0.09687 13.14 <2e-16 ***
## FirstAuthorFemale1 0.00769 0.04023 0.19 0.848
## LastAuthorFemale1 0.03134 0.04403 0.71 0.477
## Year1997 0.20848 0.13681 1.52 0.128
## Year1998 0.07226 0.18730 0.39 0.700
## Year1999 -0.17306 0.13166 -1.31 0.189
## Year2000 -0.04164 0.12385 -0.34 0.737
## Year2001 -0.21714 0.17680 -1.23 0.220
## Year2002 -0.04202 0.14664 -0.29 0.775
## Year2003 -0.15639 0.13412 -1.17 0.244
## Year2004 0.00341 0.11708 0.03 0.977
## Year2005 -0.30603 0.15946 -1.92 0.056 .
```

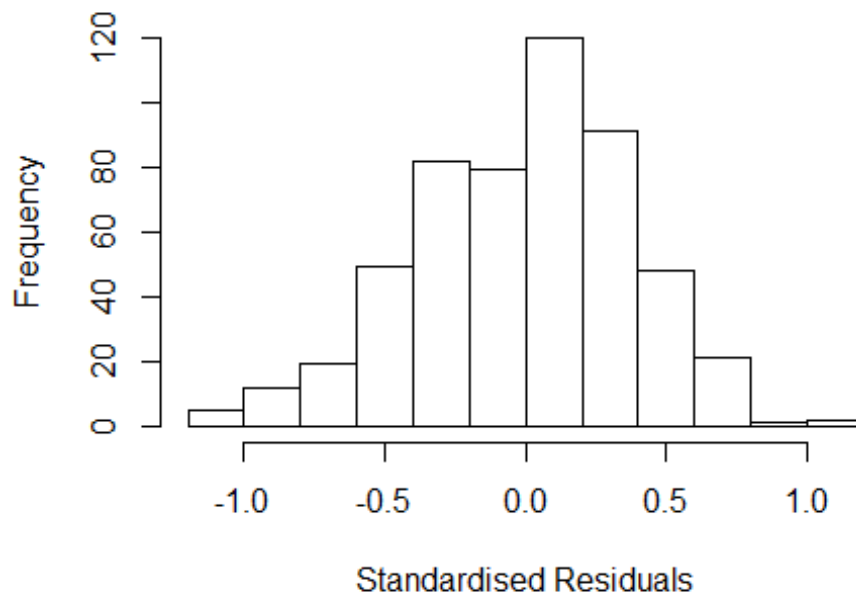


```

## Year2006          -0.16173      0.11842      -1.37      0.173
## Year2007          -0.11674      0.11637      -1.00      0.316
## Year2008          -0.04666      0.11087      -0.42      0.674
## Year2009          -0.01114      0.10673      -0.10      0.917
## Year2010          -0.10650      0.10572      -1.01      0.314
## Year2011          -0.04096      0.10850      -0.38      0.706
## Year2012          -0.07978      0.10468      -0.76      0.446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0107
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 489 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.355  0.886  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      1.89e-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 1.197 1          1.094
## Year              1.197 16          1.006

```

Residuals from first author



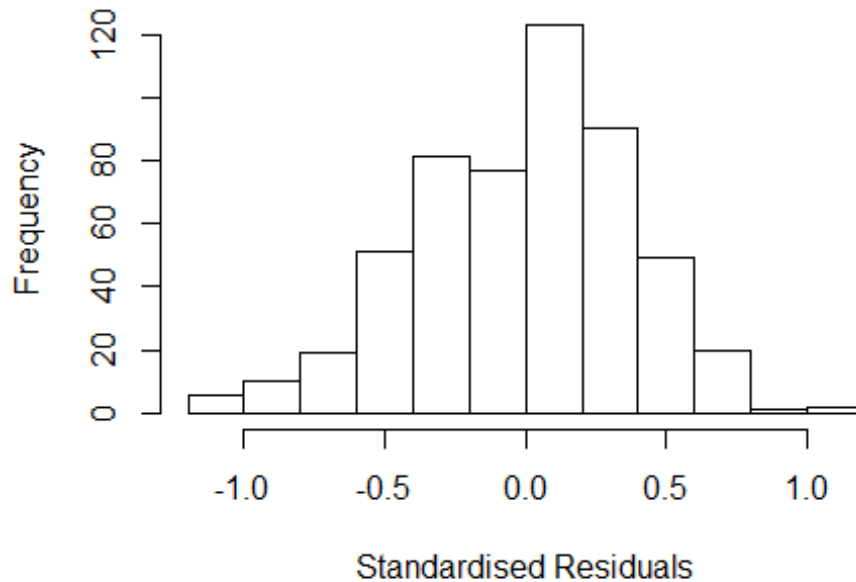
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1954 -0.2689 0.0404 0.2632 1.0400
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.277805 0.095737 13.35 <2e-16 ***
## FirstAuthorFemale1 0.011231 0.040048 0.28 0.779
## Year1997 0.211629 0.135477 1.56 0.119
## Year1998 0.077477 0.183860 0.42 0.674
## Year1999 -0.177005 0.130722 -1.35 0.176
## Year2000 -0.045278 0.123431 -0.37 0.714
## Year2001 -0.216155 0.175243 -1.23 0.218
## Year2002 -0.048290 0.145812 -0.33 0.741
## Year2003 -0.153739 0.133377 -1.15 0.250
## Year2004 0.000369 0.116825 0.00 0.997
## Year2005 -0.303886 0.158122 -1.92 0.055 .
## Year2006 -0.162207 0.118194 -1.37 0.171
```

```

## Year2007          -0.118026    0.116128    -1.02     0.310
## Year2008          -0.047205    0.110624    -0.43     0.670
## Year2009          -0.010982    0.106133    -0.10     0.918
## Year2010          -0.109273    0.105003    -1.04     0.299
## Year2011          -0.040904    0.107745    -0.38     0.704
## Year2012          -0.082405    0.103864    -0.79     0.428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0437, Adjusted R-squared:  0.0118
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 490 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.353  0.885   0.954   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      1.89e-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 1.092 1          1.045
## Year            1.092 16          1.003

```

Residuals from last author



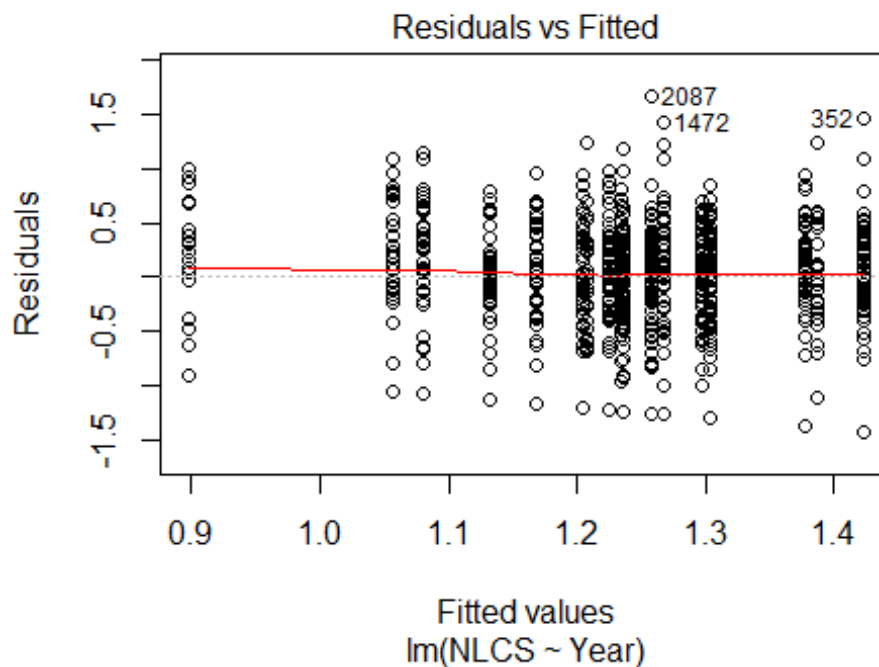
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1958 -0.2670 0.0294 0.2664 1.0341
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2747 0.0954 13.36 <2e-16 ***
## LastAuthorFemale1 0.0326 0.0439 0.74 0.457
## Year1997 0.2061 0.1356 1.52 0.129
## Year1998 0.0698 0.1866 0.37 0.709
## Year1999 -0.1739 0.1315 -1.32 0.187
## Year2000 -0.0431 0.1227 -0.35 0.726
## Year2001 -0.2157 0.1776 -1.21 0.225
## Year2002 -0.0414 0.1466 -0.28 0.778
## Year2003 -0.1565 0.1335 -1.17 0.242
## Year2004 0.0028 0.1169 0.02 0.981
## Year2005 -0.3078 0.1578 -1.95 0.052 .
## Year2006 -0.1603 0.1178 -1.36 0.174
```

```

## Year2007          -0.1166      0.1161    -1.00      0.316
## Year2008          -0.0468      0.1105    -0.42      0.672
## Year2009          -0.0108      0.1066    -0.10      0.919
## Year2010          -0.1061      0.1055    -1.01      0.315
## Year2011          -0.0397      0.1083    -0.37      0.714
## Year2012          -0.0790      0.1045    -0.76      0.450
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0126
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 486 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.350  0.885   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      1.89e-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: 529"
## [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
##   95   97  135  103   94  108  104   74  139  107  121  122  119  165  166
## 2011 2012
##  166  167
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   44   52   40   22   26   31   25   43   34   48   46   48   71   72
## 2011 2012

```

```
##      83      75
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    39   33   39   35   18   19   28   23   37   21   37   32   38   63   62
## 2011 2012
##    65   60
## [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
```



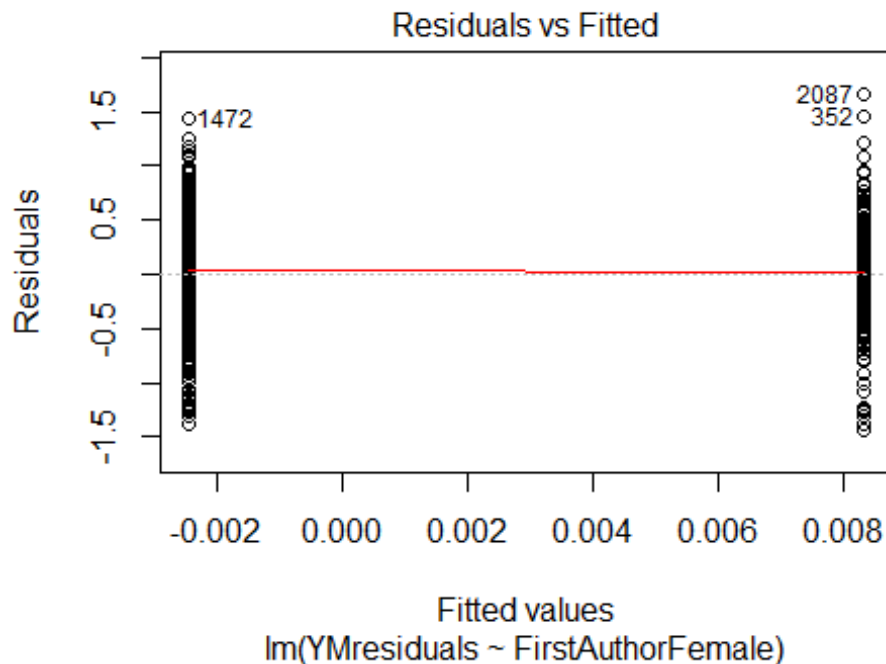
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 3.81457379461057"
## [1] "Male first author team size 2018 geometric mean: 3.32206760865827"

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

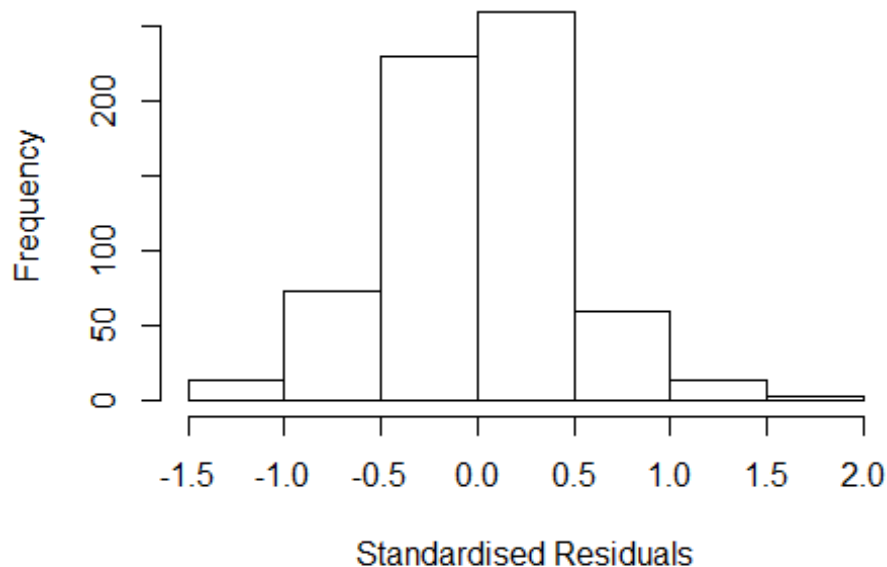
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 240, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.46410161513775"
## [1] "Male last author team size 2018 geometric mean: 3.42462058422944"

## 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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.219  1      1.104
## LastAuthorFemale  1.242  1      1.114
## UniqueAuthors    1.920  4      1.085
## Year              2.627 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
## -1.436 -0.278 0.012 0.274 1.797
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.64325 0.13048 4.93 1.1e-06 ***
## FirstAuthorFemale1 -0.05197 0.04234 -1.23 0.22010
## LastAuthorFemale1 -0.05916 0.06478 -0.91 0.36148
## UniqueAuthors2 0.45173 0.11379 3.97 8.0e-05 ***
## UniqueAuthors3 0.51404 0.10877 4.73 2.8e-06 ***
## UniqueAuthors4 0.58183 0.10842 5.37 1.1e-07 ***
## UniqueAuthors5 0.59795 0.10677 5.60 3.2e-08 ***
## Year1997 0.02095 0.10831 0.19 0.84669
## Year1998 0.28484 0.09850 2.89 0.00396 **
## Year1999 0.00403 0.09323 0.04 0.96551
```

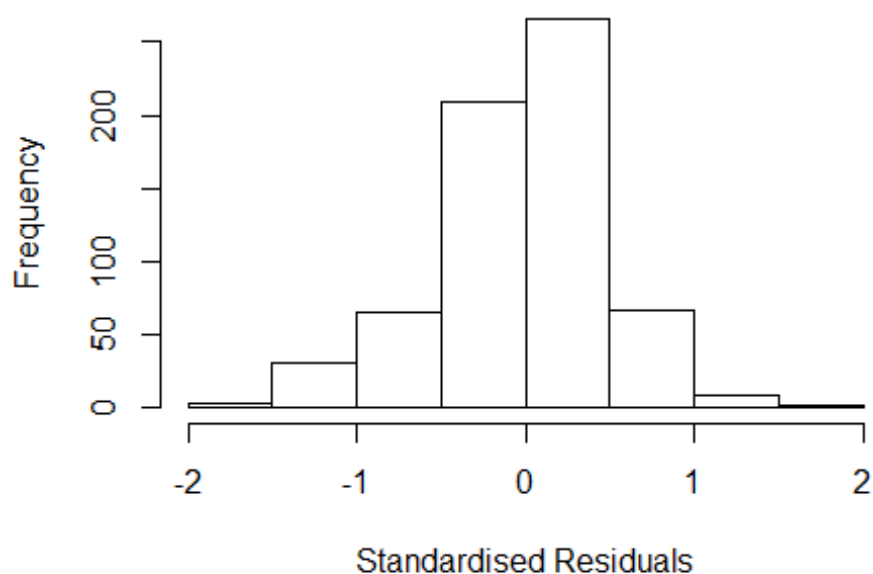


```

## Year2000          0.26375    0.15729    1.68  0.09407 .
## Year2001          0.18915    0.13968    1.35  0.17619
## Year2002          0.10331    0.11714    0.88  0.37816
## Year2003         -0.19674    0.14960   -1.32  0.18896
## Year2004         -0.01688    0.12998   -0.13  0.89675
## Year2005          0.19472    0.14284    1.36  0.17332
## Year2006          0.19858    0.10369    1.92  0.05593 .
## Year2007          0.11638    0.13400    0.87  0.38543
## Year2008          0.35502    0.09707    3.66  0.00028 ***
## Year2009          0.17475    0.08916    1.96  0.05045 .
## Year2010          0.08718    0.09259    0.94  0.34679
## Year2011          0.15603    0.08384    1.86  0.06321 .
## Year2012          0.10272    0.08773    1.17  0.24213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.186, Adjusted R-squared:  0.158
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 586 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0196 0.8550 0.9480 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      1.54e-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 1.221 1      1.105
## LastAuthorFemale  1.225 1      1.107
## Year              1.492 16      1.013

```

Residuals from first and last author



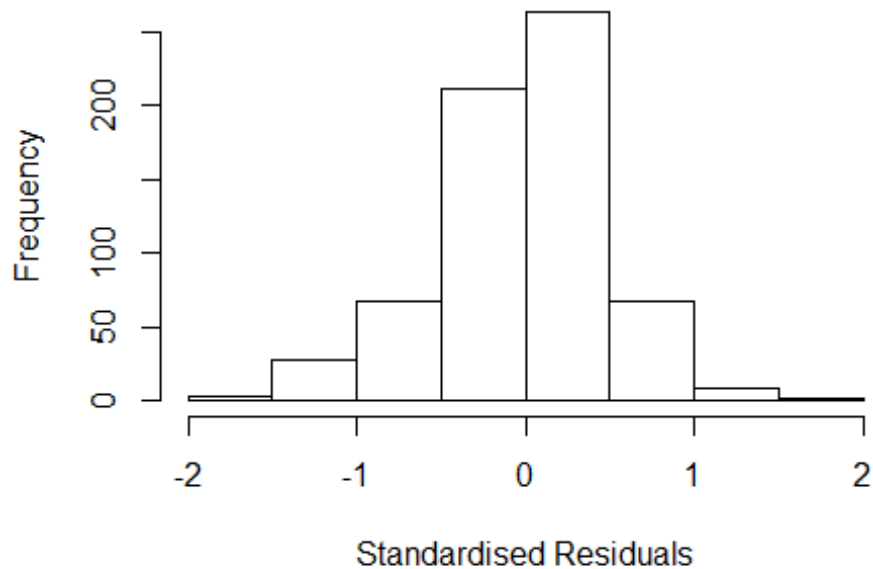
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5362 -0.2997 0.0195 0.2855 1.6596
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18010 0.07342 16.07 < 2e-16 ***
## FirstAuthorFemale1 -0.00141 0.04499 -0.03 0.97501
## LastAuthorFemale1 -0.06044 0.07526 -0.80 0.42225
## Year1997 -0.01799 0.12172 -0.15 0.88254
## Year1998 0.17156 0.09791 1.75 0.08022 .
## Year1999 -0.03525 0.09548 -0.37 0.71207
## Year2000 0.13070 0.14813 0.88 0.37795
## Year2001 0.10195 0.13060 0.78 0.43530
## Year2002 0.04291 0.11103 0.39 0.69927
## Year2003 -0.33096 0.18096 -1.83 0.06788 .
## Year2004 -0.07453 0.16216 -0.46 0.64593
## Year2005 0.09573 0.15062 0.64 0.52528
```

```

## Year2006          0.06938      0.10943      0.63  0.52633
## Year2007          0.09702      0.14865      0.65  0.51420
## Year2008          0.35611      0.10123      3.52  0.00047 ***
## Year2009          0.15927      0.08969      1.78  0.07626 .
## Year2010          0.08874      0.09186      0.97  0.33439
## Year2011          0.15604      0.08465      1.84  0.06572 .
## Year2012          0.08909      0.09017      0.99  0.32353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.067, Adjusted R-squared:  0.0404
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 588 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0942 0.8500 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      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 1.209 1      1.100
## Year              1.209 16      1.006

```

Residuals from first author



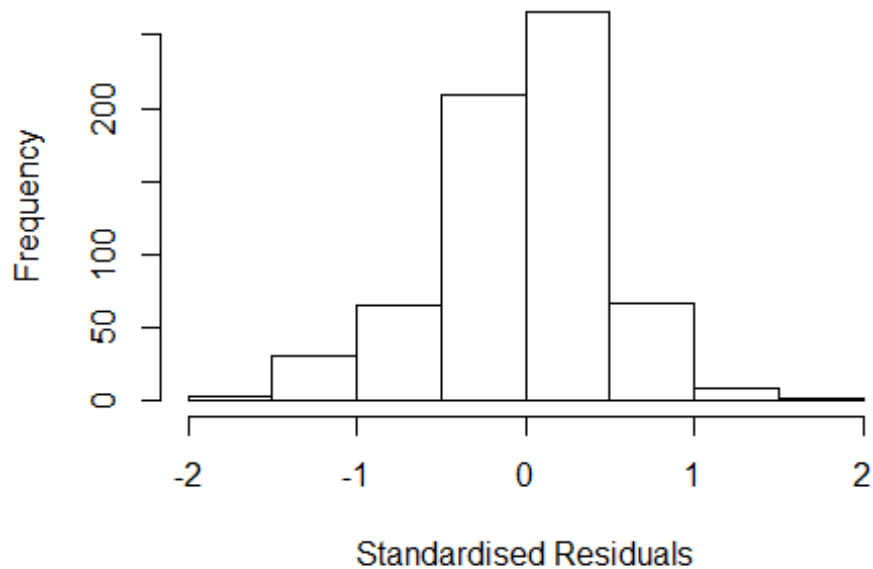
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5264 -0.2973 0.0204 0.2866 1.6635
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17494 0.07360 15.96 < 2e-16 ***
## FirstAuthorFemale1 -0.00349 0.04498 -0.08 0.93813
## Year1997 -0.01321 0.12253 -0.11 0.91419
## Year1998 0.17348 0.09881 1.76 0.07963 .
## Year1999 -0.03391 0.09578 -0.35 0.72339
## Year2000 0.12690 0.14601 0.87 0.38511
## Year2001 0.09857 0.13125 0.75 0.45293
## Year2002 0.04540 0.11031 0.41 0.68084
## Year2003 -0.32536 0.18395 -1.77 0.07741 .
## Year2004 -0.07285 0.16499 -0.44 0.65897
## Year2005 0.10153 0.14991 0.68 0.49845
## Year2006 0.07299 0.10793 0.68 0.49913
```

```

## Year2007          0.09741    0.14971    0.65  0.51550
## Year2008          0.35144    0.10130    3.47  0.00056 ***
## Year2009          0.15993    0.09031    1.77  0.07707 .
## Year2010          0.09202    0.09225    1.00  0.31893
## Year2011          0.15139    0.08518    1.78  0.07599 .
## Year2012          0.08095    0.09011    0.90  0.36937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0657, Adjusted R-squared:  0.0405
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 591 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0853 0.8500 0.9500 0.8800 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      1.54e-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 1.231 1      1.109
## Year              1.231 16      1.007

```

Residuals from last author



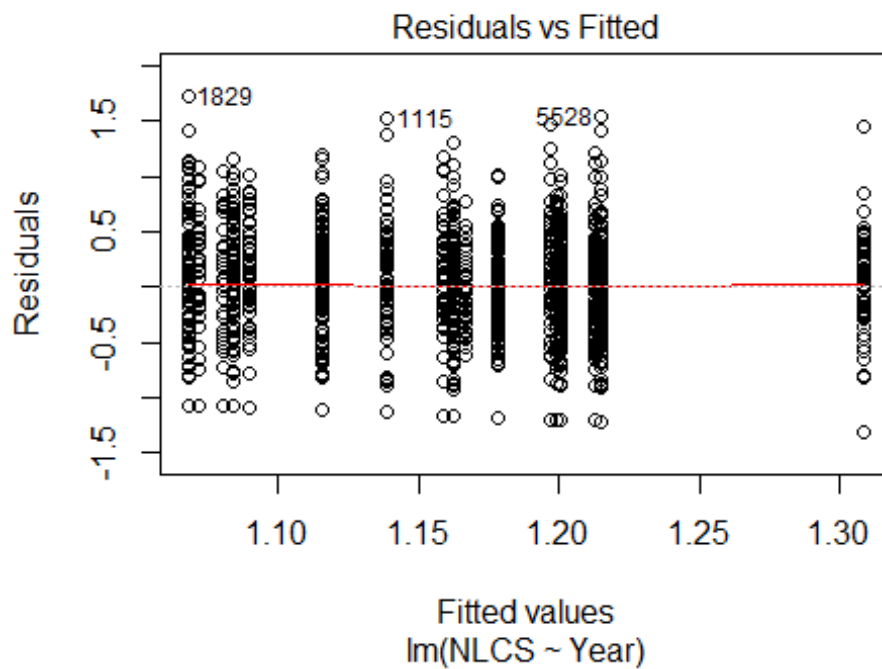
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5371 -0.3000 0.0192 0.2852 1.6582
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1797 0.0730 16.16 < 2e-16 ***
## LastAuthorFemale1 -0.0595 0.0757 -0.79 0.43220
## Year1997 -0.0168 0.1219 -0.14 0.89045
## Year1998 0.1720 0.0979 1.76 0.07939 .
## Year1999 -0.0353 0.0956 -0.37 0.71248
## Year2000 0.1303 0.1482 0.88 0.37974
## Year2001 0.1014 0.1306 0.78 0.43772
## Year2002 0.0430 0.1112 0.39 0.69925
## Year2003 -0.3308 0.1812 -1.83 0.06840 .
## Year2004 -0.0726 0.1628 -0.45 0.65595
## Year2005 0.0978 0.1505 0.65 0.51582
## Year2006 0.0701 0.1074 0.65 0.51439
```

```

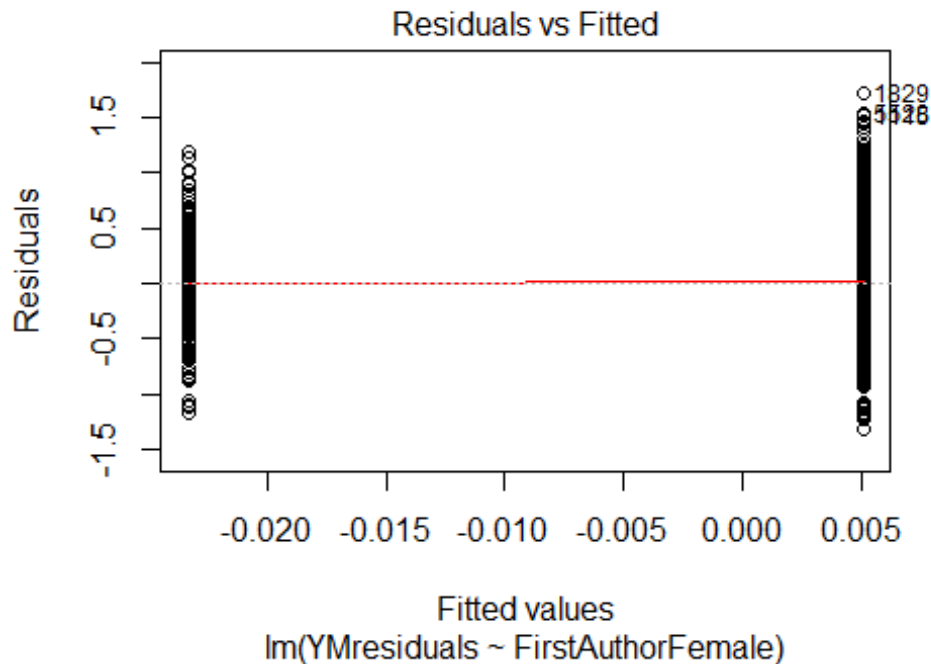
## Year2007          0.0985      0.1487      0.66  0.50818
## Year2008          0.3574      0.1014      3.52  0.00046 ***
## Year2009          0.1596      0.0895      1.78  0.07510 .
## Year2010          0.0891      0.0923      0.96  0.33493
## Year2011          0.1566      0.0847      1.85  0.06513 .
## Year2012          0.0890      0.0895      0.99  0.32033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0675, Adjusted R-squared:  0.0424
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 589 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0865 0.8480 0.9480 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      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: 649"
## [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
## 227 156 239 209 254 254 219 158 175 162 245 275 273 294 343
## 2011 2012
## 360 366
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 38 63 45 74 65 92 40 63 53 97 118 103 129 156
## 2011 2012

```

```
## 168 175
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 32 49 42 66 52 79 35 56 43 83 96 92 100 132
## 2011 2012
## 140 141
## [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
```

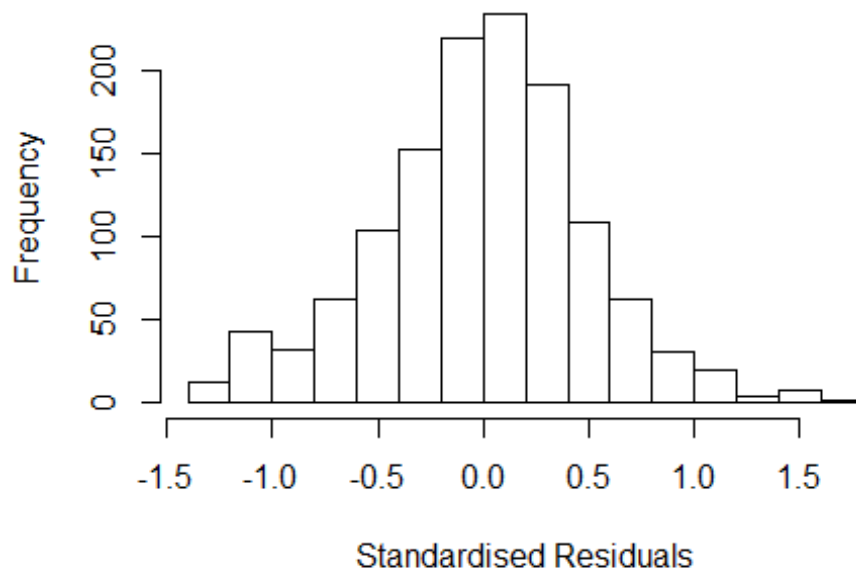


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.8, df = 1, p-value = 0.02
```

```
## [1] "Female first author team size 2018 geometric mean: 4.26485101028299"
## [1] "Male first author team size 2018 geometric mean: 3.91875894109894"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 610, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.07516600457323"
## [1] "Male last author team size 2018 geometric mean: 3.96968082293847"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 520, 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 1.065 1      1.032
## LastAuthorFemale  1.106 1      1.051
## UniqueAuthors    1.484 4      1.051
## Year              1.591 16     1.015
```

Residuals from first and last author and team size



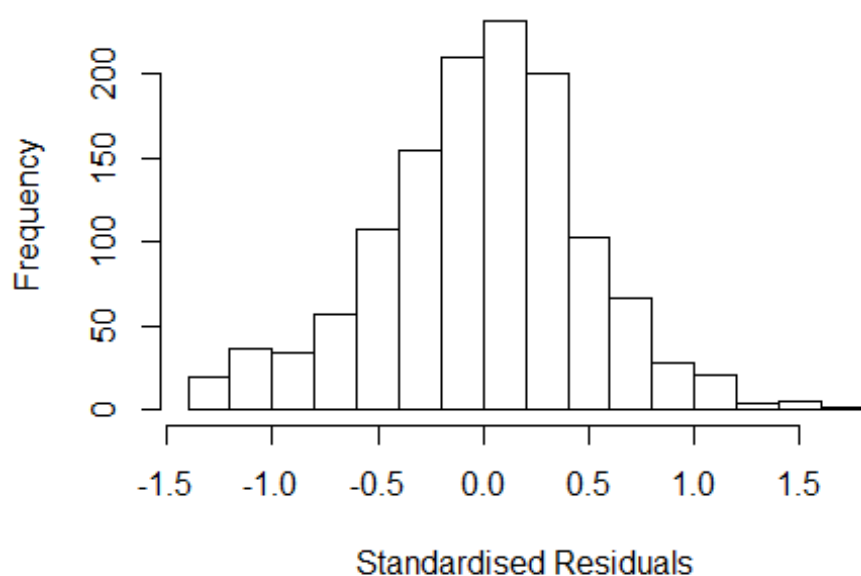
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3104 -0.2948 0.0194 0.3017 1.6779
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.31247 0.08654 15.17 <2e-16 ***
## FirstAuthorFemale1 -0.06934 0.03413 -2.03 0.0424 *
## LastAuthorFemale1 -0.00779 0.04487 -0.17 0.8622
## UniqueAuthors2 -0.01745 0.06337 -0.28 0.7831
## UniqueAuthors3 0.04069 0.06302 0.65 0.5186
## UniqueAuthors4 0.11380 0.06502 1.75 0.0803 .
## UniqueAuthors5 0.13680 0.06542 2.09 0.0367 *
## Year1997 -0.24838 0.15496 -1.60 0.1092
## Year1998 -0.21836 0.10273 -2.13 0.0337 *
## Year1999 -0.18655 0.09387 -1.99 0.0471 *
```

```

## Year2000      -0.17414    0.10222   -1.70    0.0887 .
## Year2001      -0.13890    0.14037   -0.99    0.3226
## Year2002      -0.23905    0.11868   -2.01    0.0442 *
## Year2003      -0.15323    0.12843   -1.19    0.2331
## Year2004      -0.17771    0.10383   -1.71    0.0872 .
## Year2005      -0.11769    0.10887   -1.08    0.2799
## Year2006      -0.26082    0.09954   -2.62    0.0089 **
## Year2007      -0.15573    0.08676   -1.79    0.0729 .
## Year2008      -0.14391    0.08835   -1.63    0.1036
## Year2009      -0.18863    0.09036   -2.09    0.0370 *
## Year2010      -0.25809    0.08798   -2.93    0.0034 **
## Year2011      -0.15389    0.08399   -1.83    0.0672 .
## Year2012      -0.16678    0.08693   -1.92    0.0553 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.0297, Adjusted R-squared:  0.0127
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1174 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.854  0.952  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      7.82e-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 1.063 1      1.031
## LastAuthorFemale  1.068 1      1.033
## Year              1.120 16      1.004

```

Residuals from first and last author



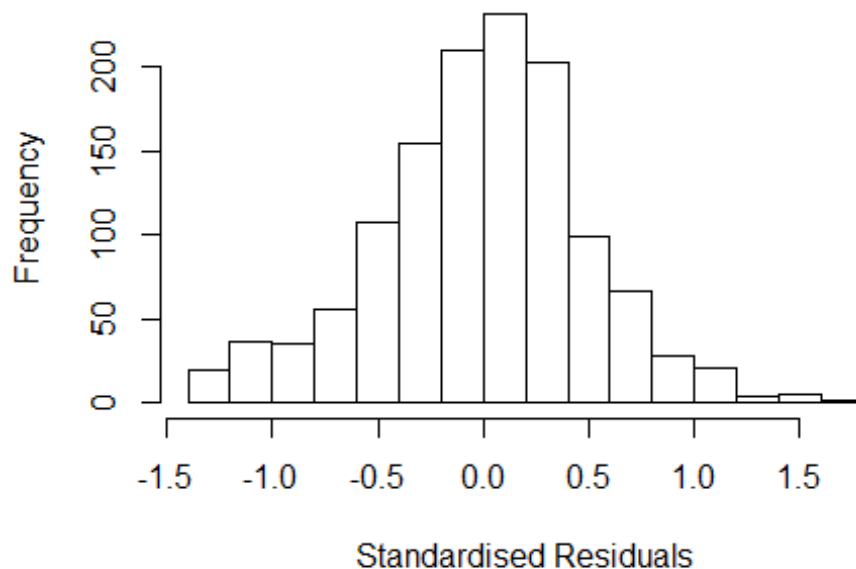
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3503 -0.2993  0.0236  0.3037  1.6752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3448    0.0731   18.39  <2e-16 ***
## FirstAuthorFemale1 -0.0586    0.0343   -1.71   0.0882 .
## LastAuthorFemale1  0.0055    0.0442    0.12   0.9009
## Year1997          -0.2483    0.1535   -1.62   0.1060
## Year1998          -0.2109    0.1023   -2.06   0.0394 *
## Year1999          -0.1732    0.0897   -1.93   0.0536 .
## Year2000          -0.1876    0.1006   -1.87   0.0623 .
## Year2001          -0.1386    0.1315   -1.05   0.2917
## Year2002          -0.2280    0.1179   -1.93   0.0534 .
## Year2003          -0.1436    0.1244   -1.15   0.2485
## Year2004          -0.1616    0.1022   -1.58   0.1142
## Year2005          -0.0941    0.1069   -0.88   0.3787
```

```

## Year2006          -0.2521      0.0966   -2.61   0.0091 **
## Year2007          -0.1375      0.0838   -1.64   0.1012
## Year2008          -0.1187      0.0857   -1.38   0.1663
## Year2009          -0.1636      0.0872   -1.88   0.0607 .
## Year2010          -0.2188      0.0834   -2.62   0.0088 **
## Year2011          -0.1181      0.0799   -1.48   0.1397
## Year2012          -0.1271      0.0827   -1.54   0.1244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.000854
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1169 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.153  0.863  0.951  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      7.82e-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 1.056 1      1.028
## Year      1.056 16      1.002

```

Residuals from first author



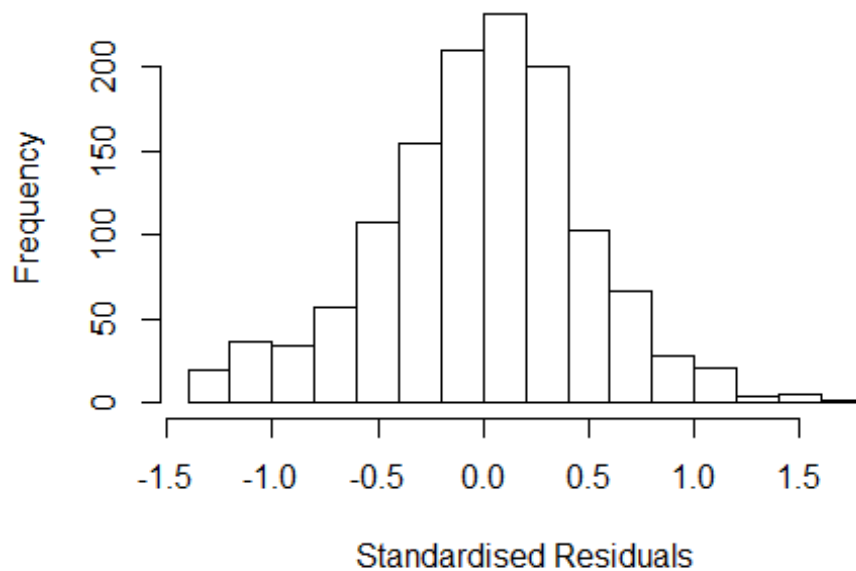
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3448 -0.2983  0.0242  0.3040  1.6749
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3448    0.0732   18.38  <2e-16 ***
## FirstAuthorFemale1 -0.0584    0.0343   -1.71   0.0884 .
## Year1997         -0.2478    0.1533   -1.62   0.1064
## Year1998         -0.2102    0.1021   -2.06   0.0396 *
## Year1999         -0.1728    0.0894   -1.93   0.0535 .
## Year2000         -0.1873    0.1006   -1.86   0.0630 .
## Year2001         -0.1381    0.1311   -1.05   0.2926
## Year2002         -0.2278    0.1180   -1.93   0.0538 .
## Year2003         -0.1428    0.1239   -1.15   0.2493
## Year2004         -0.1612    0.1022   -1.58   0.1149
## Year2005         -0.0936    0.1067   -0.88   0.3805
## Year2006         -0.2517    0.0966   -2.61   0.0093 **
```

```

## Year2007          -0.1364      0.0831   -1.64   0.1009
## Year2008          -0.1184      0.0856   -1.38   0.1671
## Year2009          -0.1633      0.0870   -1.88   0.0608 .
## Year2010          -0.2183      0.0833   -2.62   0.0089 **
## Year2011          -0.1175      0.0797   -1.47   0.1407
## Year2012          -0.1262      0.0821   -1.54   0.1246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.00163
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1168 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.153  0.863  0.951  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      7.82e-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 1.061 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
## -1.3399 -0.3062  0.0146  0.3090  1.6847
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.33705    0.07258   18.42  <2e-16 ***
## LastAuthorFemale1 0.00289    0.04415    0.07  0.9478
## Year1997       -0.24467    0.15194   -1.61  0.1076
## Year1998       -0.20718    0.10168   -2.04  0.0418 *
## Year1999       -0.17573    0.08982   -1.96  0.0506 .
## Year2000       -0.18730    0.10032   -1.87  0.0621 .
## Year2001       -0.13431    0.13072   -1.03  0.3044
## Year2002       -0.22980    0.11794   -1.95  0.0516 .
## Year2003       -0.15327    0.12458   -1.23  0.2188
## Year2004       -0.16165    0.10205   -1.58  0.1134
## Year2005       -0.09500    0.10701   -0.89  0.3749
## Year2006       -0.25457    0.09614   -2.65  0.0082 **
```

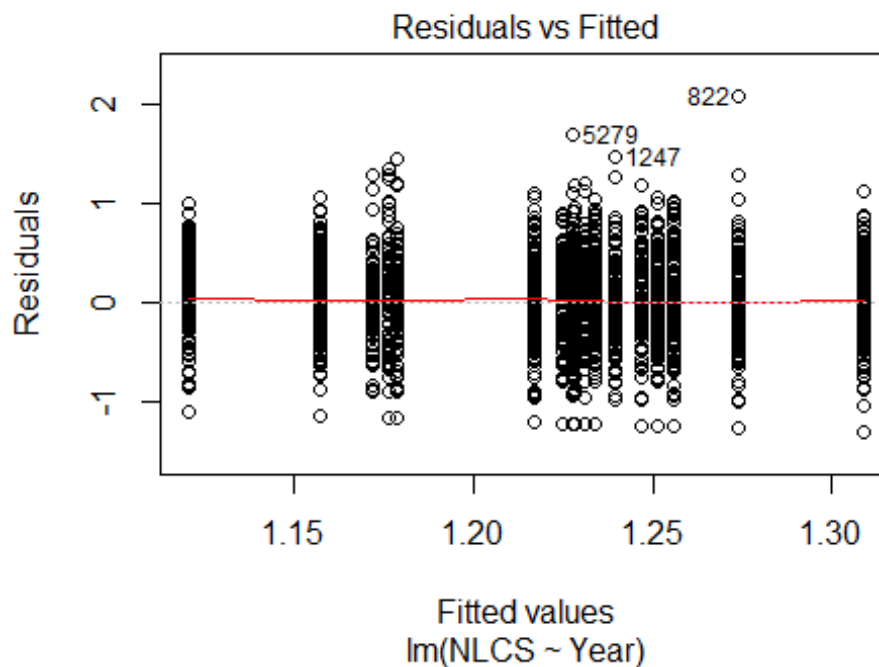


```

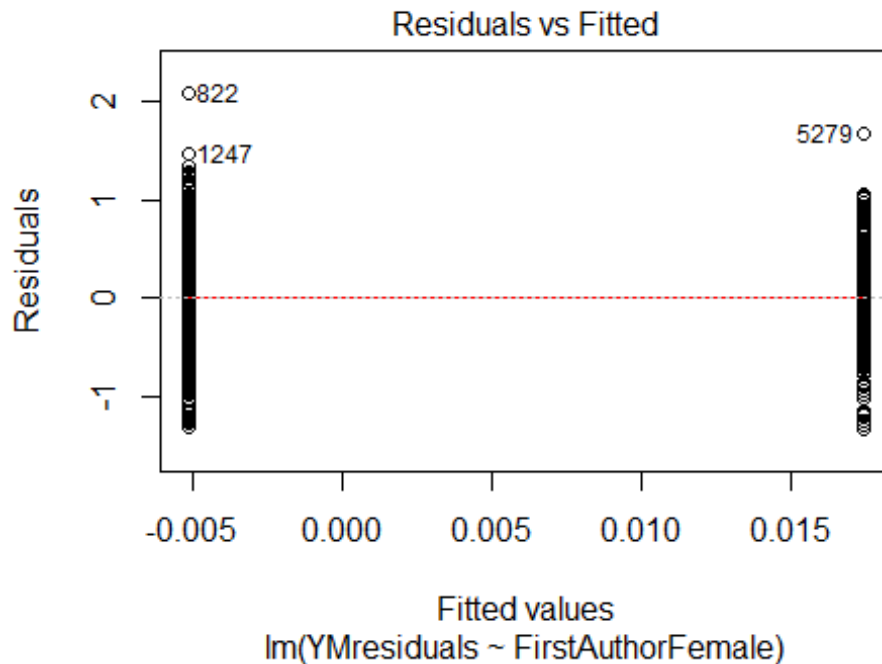
## Year2007          -0.14203      0.08361    -1.70    0.0896 .
## Year2008          -0.12243      0.08540    -1.43    0.1519
## Year2009          -0.17050      0.08704    -1.96    0.0503 .
## Year2010          -0.22632      0.08363    -2.71    0.0069 **
## Year2011          -0.12178      0.07990    -1.52    0.1277
## Year2012          -0.13198      0.08261    -1.60    0.1104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0126, Adjusted R-squared:  -0.000675
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1161 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.152  0.862  0.951  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      7.82e-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: 1278"
## [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
##  304  279  292  255  286  277  227  234  266  288  309  328  309  402  420
## 2011 2012
##  392  362
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  130  116  129  121  113   94  128   98  123  134  152  183  155  174  195
## 2011 2012

```

```
## 193 176
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 98 113 107 102 80 118 86 103 115 134 153 139 149 163
## 2011 2012
## 168 154
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



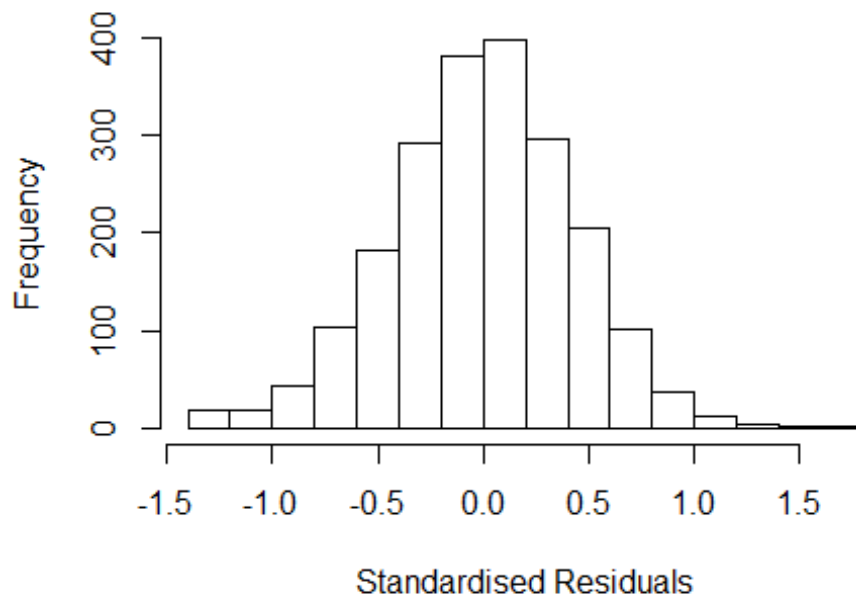
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0087, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 4.43168253632928"
## [1] "Male first author team size 2018 geometric mean: 4.03149637023736"
## 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 = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.71300338254507"
## [1] "Male last author team size 2018 geometric mean: 3.97275364767105"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, 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	1.076	1	1.037
LastAuthorFemale	1.059	1	1.029
UniqueAuthors	1.295	4	1.033
Year	1.328	16	1.009

Residuals from first and last author and team size



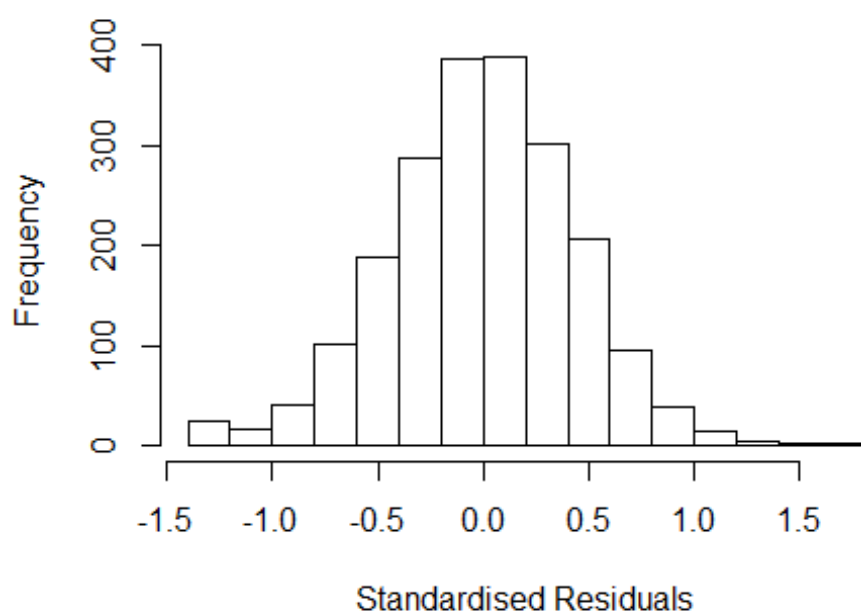
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31570 -0.27463  0.00121  0.28255  1.69507
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16077    0.07356   15.78  <2e-16 ***
## FirstAuthorFemale1  0.02817    0.02359    1.19   0.233
## LastAuthorFemale1  0.01716    0.02990    0.57   0.566
## UniqueAuthors2     0.06294    0.06313    1.00   0.319
## UniqueAuthors3     0.07111    0.06251    1.14   0.255
## UniqueAuthors4     0.06744    0.06211    1.09   0.278
## UniqueAuthors5     0.10447    0.06274    1.67   0.096 .
## Year1997        -0.00854    0.06442   -0.13   0.895
## Year1998         0.04018    0.06165    0.65   0.515
## Year1999        -0.04390    0.06008   -0.73   0.465
```

```

## Year2000      -0.03491    0.05948   -0.59    0.557
## Year2001      -0.05936    0.06444   -0.92    0.357
## Year2002       0.04949    0.06524    0.76    0.448
## Year2003      -0.06645    0.07172   -0.93    0.354
## Year2004      -0.08660    0.06508   -1.33    0.183
## Year2005       0.02747    0.05870    0.47    0.640
## Year2006       0.02224    0.05831    0.38    0.703
## Year2007      -0.08508    0.05818   -1.46    0.144
## Year2008       0.08381    0.05902    1.42    0.156
## Year2009       0.00531    0.05866    0.09    0.928
## Year2010      -0.01995    0.05898   -0.34    0.735
## Year2011      -0.00352    0.05444   -0.06    0.948
## Year2012       0.00525    0.05523    0.10    0.924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00552
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 1925 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0638 0.8660 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      4.78e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1      1.026
## LastAuthorFemale  1.039 1      1.019
## Year              1.067 16      1.002

```

Residuals from first and last author

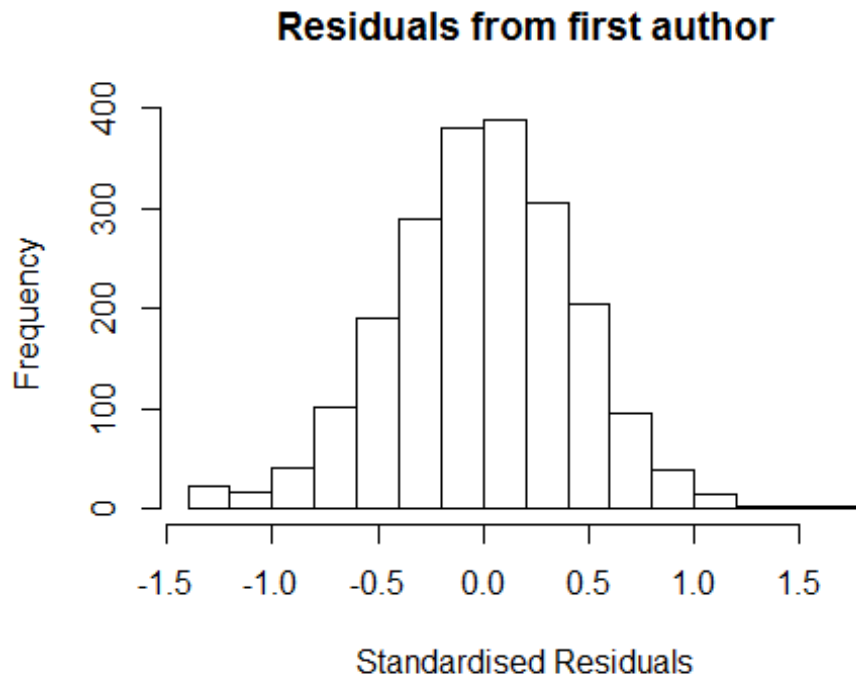


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32170 -0.27325  0.00198  0.27712  1.67423
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.22895    0.04510   27.25  <2e-16 ***
## FirstAuthorFemale1  0.03209    0.02335    1.37    0.17
## LastAuthorFemale1  0.01660    0.02966    0.56    0.58
## Year1997          -0.00445    0.06414   -0.07    0.94
## Year1998           0.03394    0.06121    0.55    0.58
## Year1999          -0.04208    0.05969   -0.70    0.48
## Year2000          -0.03680    0.05895   -0.62    0.53
## Year2001          -0.06444    0.06387   -1.01    0.31
## Year2002           0.05574    0.06478    0.86    0.39
## Year2003          -0.06066    0.07071   -0.86    0.39
## Year2004          -0.08012    0.06458   -1.24    0.21
## Year2005           0.03155    0.05849    0.54    0.59
```

```

## Year2006          0.02638    0.05825    0.45    0.65
## Year2007         -0.07738    0.05800   -1.33    0.18
## Year2008          0.09275    0.05855    1.58    0.11
## Year2009          0.01382    0.05811    0.24    0.81
## Year2010         -0.00827    0.05826   -0.14    0.89
## Year2011          0.00689    0.05409    0.13    0.90
## Year2012          0.01719    0.05464    0.31    0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.00437
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 1925 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0731 0.8670 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      4.78e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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

```



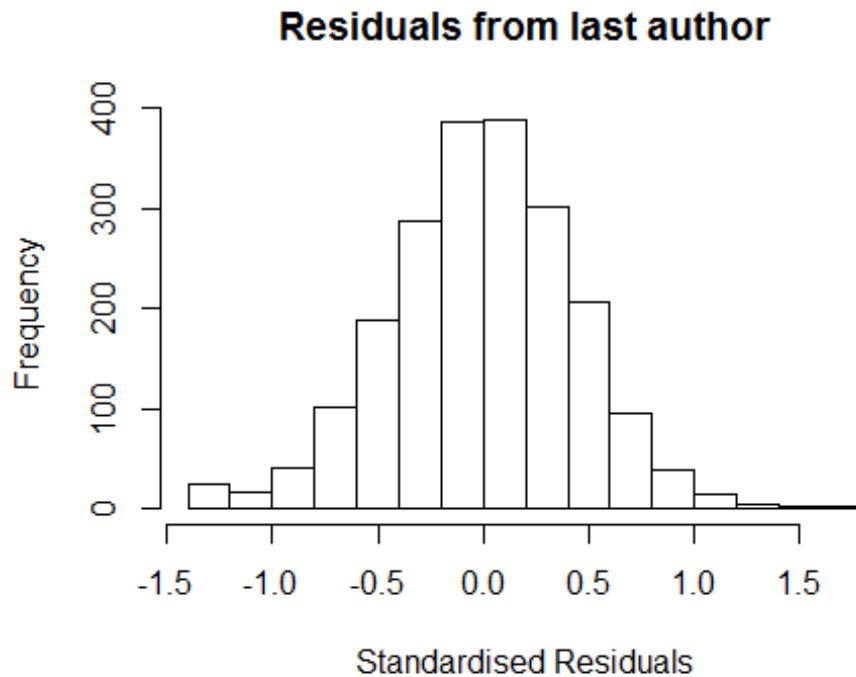
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32400 -0.27347  0.00297  0.27512  1.67076
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22955    0.04510   27.26  <2e-16 ***
## FirstAuthorFemale1 0.03354    0.02323    1.44    0.15
## Year1997      -0.00419    0.06412   -0.07    0.95
## Year1998       0.03514    0.06115    0.57    0.57
## Year1999      -0.04164    0.05969   -0.70    0.49
## Year2000      -0.03593    0.05893   -0.61    0.54
## Year2001      -0.06383    0.06381   -1.00    0.32
## Year2002       0.05727    0.06467    0.89    0.38
## Year2003      -0.05938    0.07065   -0.84    0.40
## Year2004      -0.07894    0.06461   -1.22    0.22
## Year2005       0.03361    0.05833    0.58    0.56
## Year2006       0.02794    0.05825    0.48    0.63
```



```

## Year2007          -0.07650    0.05805   -1.32    0.19
## Year2008          0.09445    0.05842    1.62    0.11
## Year2009          0.01513    0.05799    0.26    0.79
## Year2010         -0.00686    0.05822   -0.12    0.91
## Year2011          0.00878    0.05390    0.16    0.87
## Year2012          0.01881    0.05460    0.34    0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.0128, Adjusted R-squared:  0.00469
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~= 1. The remaining 1932 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0755 0.8680 0.9520 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      4.78e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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

```



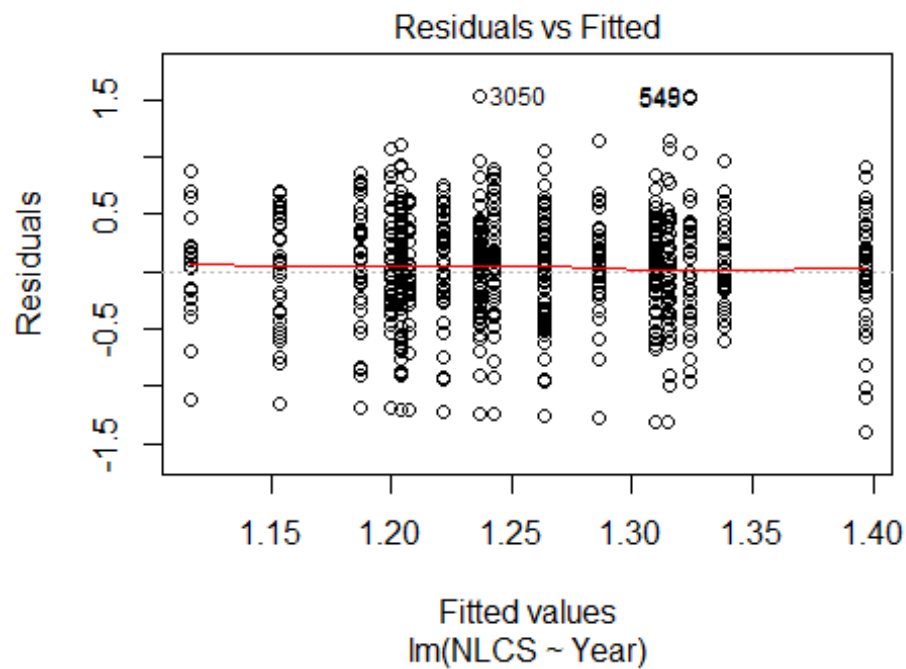
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.328960 -0.272336 -0.000794  0.280180  1.698904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.23429    0.04476   27.57  <2e-16 ***
## LastAuthorFemale1  0.02094    0.02962    0.71    0.48
## Year1997         -0.00435    0.06400   -0.07    0.95
## Year1998          0.03434    0.06119    0.56    0.57
## Year1999         -0.03874    0.05960   -0.65    0.52
## Year2000         -0.03651    0.05910   -0.62    0.54
## Year2001         -0.06439    0.06384   -1.01    0.31
## Year2002          0.05496    0.06486    0.85    0.40
## Year2003         -0.05981    0.07069   -0.85    0.40
## Year2004         -0.07880    0.06473   -1.22    0.22
## Year2005          0.03127    0.05850    0.53    0.59
## Year2006          0.03193    0.05849    0.55    0.59
```

```

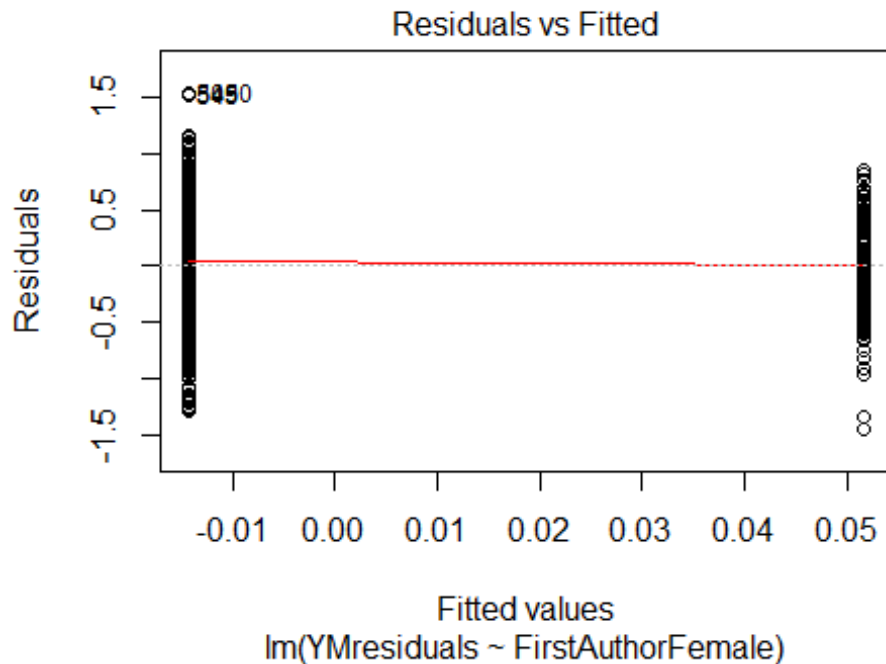
## Year2007          -0.07350      0.05811      -1.26      0.21
## Year2008           0.09467      0.05857       1.62      0.11
## Year2009           0.01719      0.05817       0.30      0.77
## Year2010          -0.00619      0.05829      -0.11      0.92
## Year2011           0.00727      0.05409       0.13      0.89
## Year2012           0.01763      0.05467       0.32      0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00394
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 175 weights are ~= 1. The remaining 1919 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0621 0.8680 0.9520 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      4.78e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2094"
## [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
## 123 106 140 118 163 148 108 85 114 139 133 114 156 210 217
## 2011 2012
## 212 182
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 36 37 37 38 42 27 35 23 42 34 43 32 48 71 77
## 2011 2012

```

```
##      83    95
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   35   35   32   37   41   26   32   20   30   28   39   27   39   60   63
## 2011 2012
##   68   77
## [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
```

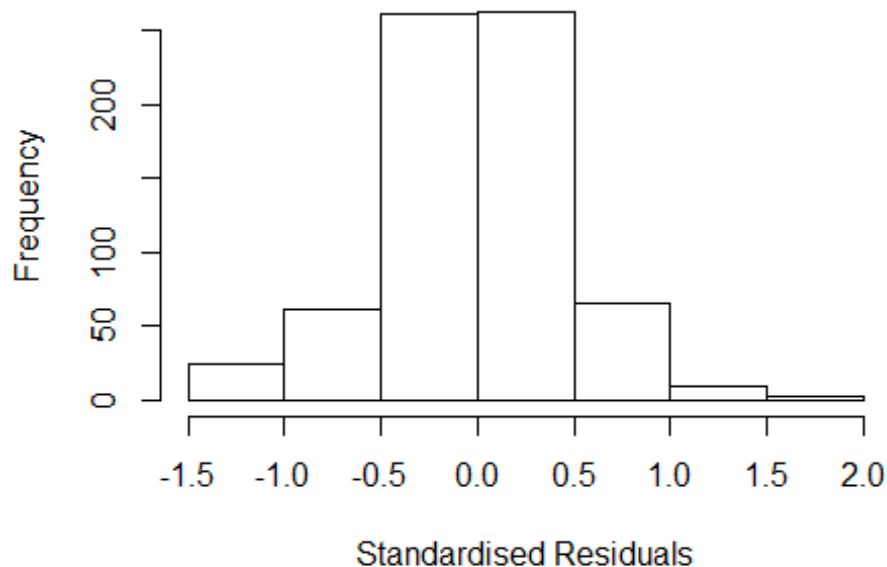


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 3e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.92556482553286"
## [1] "Male first author team size 2018 geometric mean: 3.55384126001811"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 590, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.778533119429"
## [1] "Male last author team size 2018 geometric mean: 3.66140430358699"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 480, 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 1.241 1      1.114
## LastAuthorFemale  1.255 1      1.120
## UniqueAuthors     2.520 4      1.122
## Year              3.045 16     1.035
```

Residuals from first and last author and team size



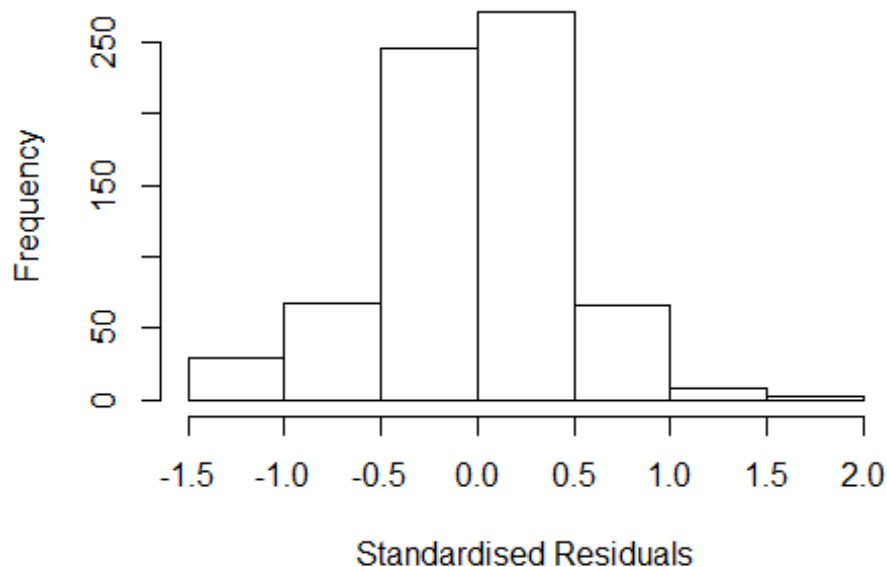
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4422 -0.2554 -0.0081  0.2646  1.6538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8681    0.1392   6.23   8e-10 ***
## FirstAuthorFemale1 0.0197    0.0385   0.51  0.6084
## LastAuthorFemale1 0.0832    0.0546   1.52  0.1283
## UniqueAuthors2    0.2713    0.1183   2.29  0.0221 *
## UniqueAuthors3    0.3229    0.1157   2.79  0.0054 **
## UniqueAuthors4    0.3560    0.1154   3.09  0.0021 **
## UniqueAuthors5    0.3051    0.1206   2.53  0.0116 *
## Year1997          0.0173    0.1168   0.15  0.8821
## Year1998          0.2022    0.0960   2.11  0.0356 *
## Year1999          0.0557    0.1108   0.50  0.6150
```

```

## Year2000          0.1623      0.1009      1.61      0.1081
## Year2001          0.1713      0.1384      1.24      0.2164
## Year2002          0.1528      0.1292      1.18      0.2371
## Year2003         -0.0199      0.1245     -0.16      0.8733
## Year2004          0.2830      0.1060      2.67      0.0078 **
## Year2005          0.0613      0.1189      0.52      0.6062
## Year2006          0.1582      0.1094      1.45      0.1484
## Year2007          0.1221      0.0991      1.23      0.2182
## Year2008          0.1484      0.0957      1.55      0.1216
## Year2009          0.1379      0.0912      1.51      0.1310
## Year2010          0.0287      0.0917      0.31      0.7540
## Year2011          0.1626      0.0763      2.13      0.0334 *
## Year2012          0.0885      0.0790      1.12      0.2628
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0772, Adjusted R-squared:  0.0468
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 630 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.041  0.852  0.950   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      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"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.223 1      1.106
## LastAuthorFemale  1.202 1      1.096
## Year              1.322 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
## -1.45525 -0.26417 0.00425 0.25957 1.62680
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19116 0.06397 18.62 <2e-16 ***
## FirstAuthorFemale1 0.02332 0.03864 0.60 0.546
## LastAuthorFemale1 0.06544 0.05188 1.26 0.208
## Year1997 0.00024 0.11538 0.00 0.998
## Year1998 0.11347 0.08548 1.33 0.185
## Year1999 0.03104 0.11007 0.28 0.778
## Year2000 0.10102 0.09824 1.03 0.304
## Year2001 0.11848 0.13105 0.90 0.366
## Year2002 0.13939 0.13082 1.07 0.287
## Year2003 -0.06008 0.11808 -0.51 0.611
## Year2004 0.24077 0.10191 2.36 0.018 *
## Year2005 -0.01178 0.12423 -0.09 0.924
```

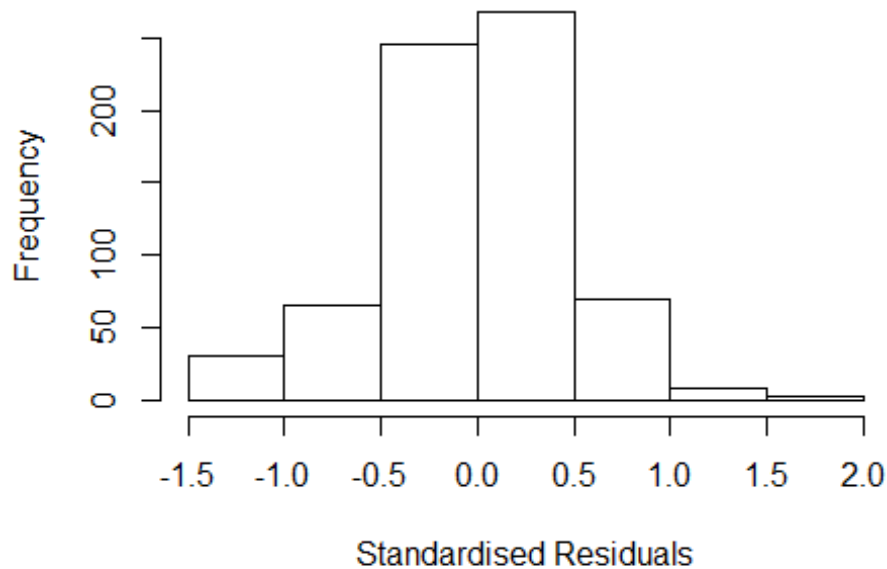


```

## Year2006          0.13644      0.10806      1.26      0.207
## Year2007          0.10191      0.09626      1.06      0.290
## Year2008          0.12923      0.09445      1.37      0.172
## Year2009          0.11606      0.08758      1.33      0.186
## Year2010          0.01927      0.09057      0.21      0.832
## Year2011          0.15869      0.07124      2.23      0.026 *
## Year2012          0.07737      0.07545      1.03      0.306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.00183
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 618 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0474 0.8420 0.9460 0.8730 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      1.45e-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 1.154 1      1.074
## Year      1.154 16      1.004

```

Residuals from first author



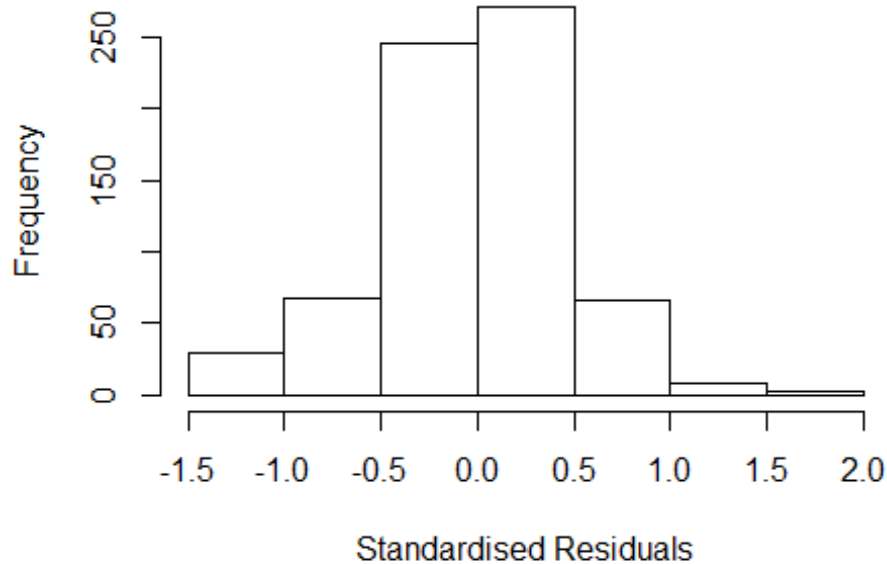
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46800 -0.26171  0.00318  0.25706  1.62163
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.19680    0.06320   18.94  <2e-16 ***
## FirstAuthorFemale1  0.03213    0.03791    0.85   0.397
## Year1997        -0.00623    0.11334   -0.05   0.956
## Year1998         0.10978    0.08535    1.29   0.199
## Year1999         0.03057    0.10966    0.28   0.781
## Year2000         0.10508    0.09848    1.07   0.286
## Year2001         0.11695    0.13015    0.90   0.369
## Year2002         0.13653    0.13133    1.04   0.299
## Year2003        -0.06284    0.11951   -0.53   0.599
## Year2004         0.23907    0.10161    2.35   0.019 *
## Year2005        -0.00642    0.12351   -0.05   0.959
## Year2006         0.13214    0.10842    1.22   0.223
```

```

## Year2007          0.10560      0.09554      1.11      0.269
## Year2008          0.12557      0.09409      1.33      0.182
## Year2009          0.11211      0.08704      1.29      0.198
## Year2010          0.01614      0.08978      0.18      0.857
## Year2011          0.16202      0.07114      2.28      0.023 *
## Year2012          0.07978      0.07529      1.06      0.290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.025, Adjusted R-squared:  0.000311
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 621 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0629 0.8440 0.9500 0.8780 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      1.45e-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 1.14 1      1.068
## Year      1.14 16      1.004

```

Residuals from last author



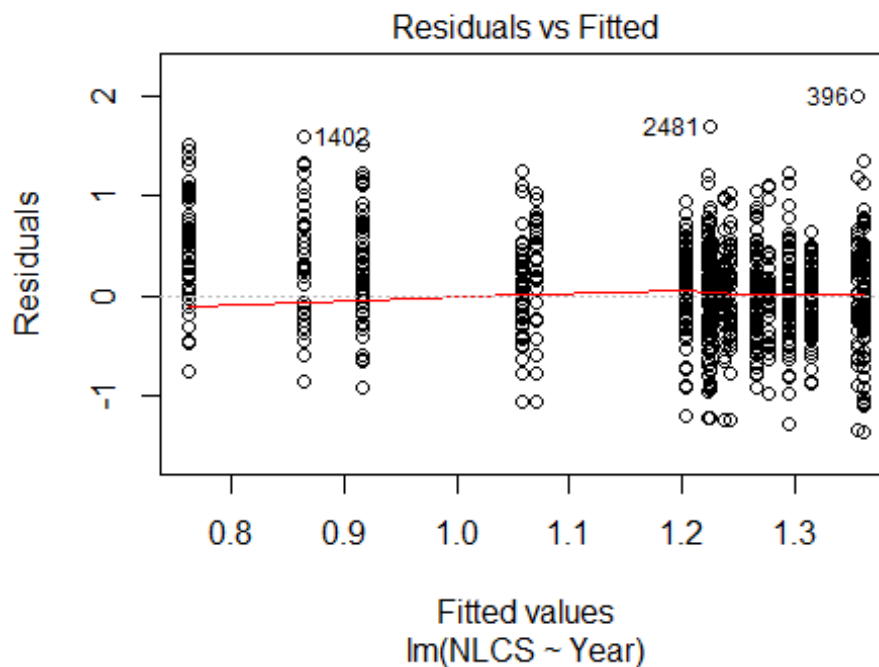
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43259 -0.25834 0.00164 0.26131 1.61907
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19376 0.06382 18.71 <2e-16 ***
## LastAuthorFemale1 0.06997 0.05073 1.38 0.168
## Year1997 0.00314 0.11504 0.03 0.978
## Year1998 0.11432 0.08554 1.34 0.182
## Year1999 0.03617 0.10977 0.33 0.742
## Year2000 0.10253 0.09819 1.04 0.297
## Year2001 0.11835 0.13156 0.90 0.369
## Year2002 0.13758 0.13066 1.05 0.293
## Year2003 -0.05274 0.11598 -0.45 0.649
## Year2004 0.23883 0.10197 2.34 0.019 *
## Year2005 -0.00904 0.12459 -0.07 0.942
## Year2006 0.14413 0.10624 1.36 0.175
```

```

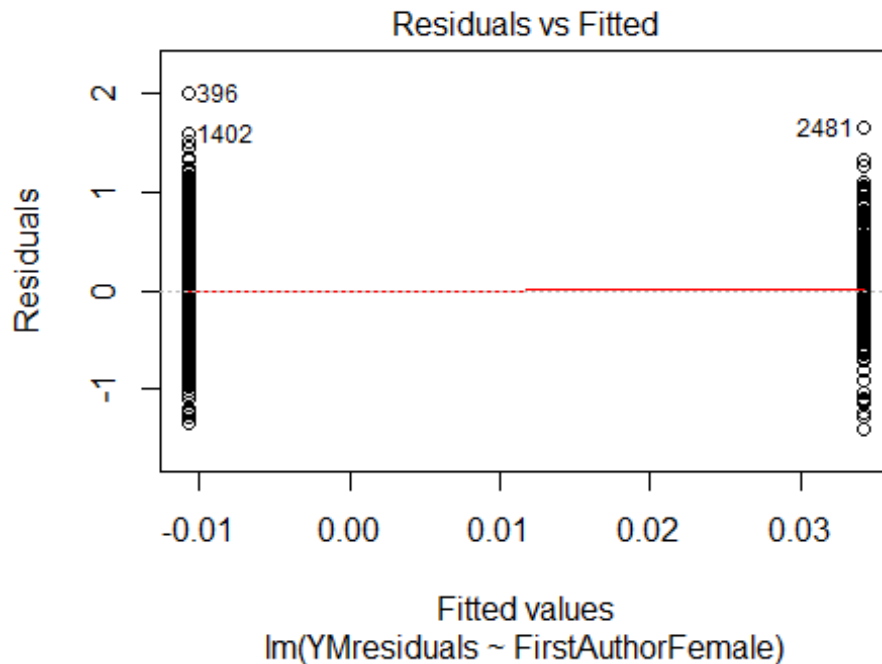
## Year2007      0.10690      0.09508      1.12      0.261
## Year2008      0.13260      0.09402      1.41      0.159
## Year2009      0.11918      0.08676      1.37      0.170
## Year2010      0.02039      0.09073      0.22      0.822
## Year2011      0.16110      0.07136      2.26      0.024 *
## Year2012      0.07945      0.07532      1.05      0.292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0272, Adjusted R-squared:  0.00258
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 614 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0537 0.8430 0.9450 0.8730 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      1.45e-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: 689"
## [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
## 167 141 131 133 139 155 237 124 131 134 126 127 119 153 189
## 2011 2012
## 151 184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 44 44 47 32 48 89 53 48 62 56 68 60 83 92
## 2011 2012

```

```
##      83      85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    51   32   38   40   28   39   81   47   40   52   52   60   52   70   78
## 2011 2012
##    68   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 = 79, df = 16, p-value = 3e-10
```



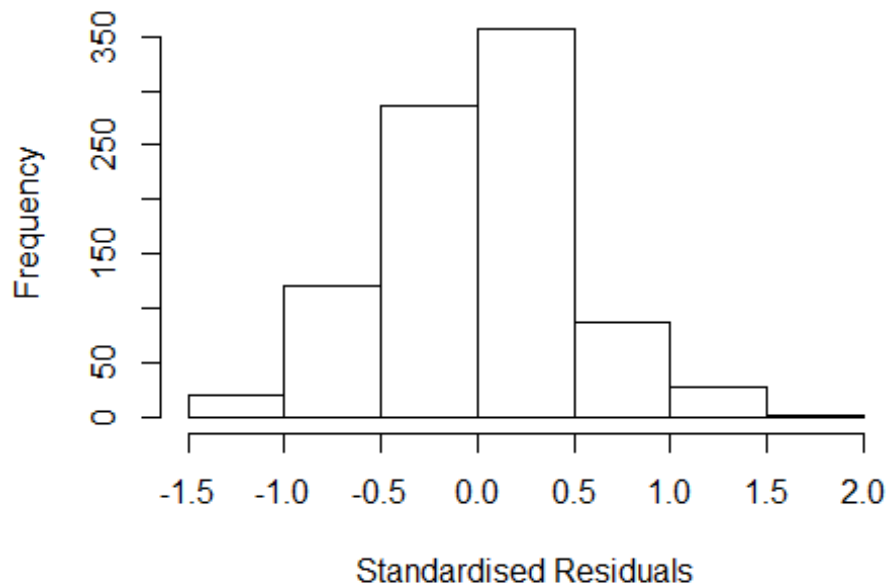
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.64, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 3.60972788570434"
## [1] "Male first author team size 2018 geometric mean: 4.24193573626398"
## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.15129477786792"
## [1] "Male last author team size 2018 geometric mean: 4.00861799395015"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 230, 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	1.190	1	1.091
LastAuthorFemale	1.158	1	1.076
UniqueAuthors	2.691	4	1.132
Year	3.017	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
## -1.4599 -0.3355 0.0311 0.3041 1.6900
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5831 0.1173 4.97 8.1e-07 ***
## FirstAuthorFemale1 0.0164 0.0386 0.42 0.671
## LastAuthorFemale1 -0.0251 0.0526 -0.48 0.633
## UniqueAuthors2 0.5614 0.0914 6.14 1.2e-09 ***
## UniqueAuthors3 0.6312 0.0843 7.49 1.7e-13 ***
## UniqueAuthors4 0.6589 0.0844 7.80 1.7e-14 ***
## UniqueAuthors5 0.6785 0.0816 8.32 3.4e-16 ***
## Year1997 0.2117 0.1492 1.42 0.156
## Year1998 0.1421 0.1362 1.04 0.297
## Year1999 0.1289 0.1447 0.89 0.373
```

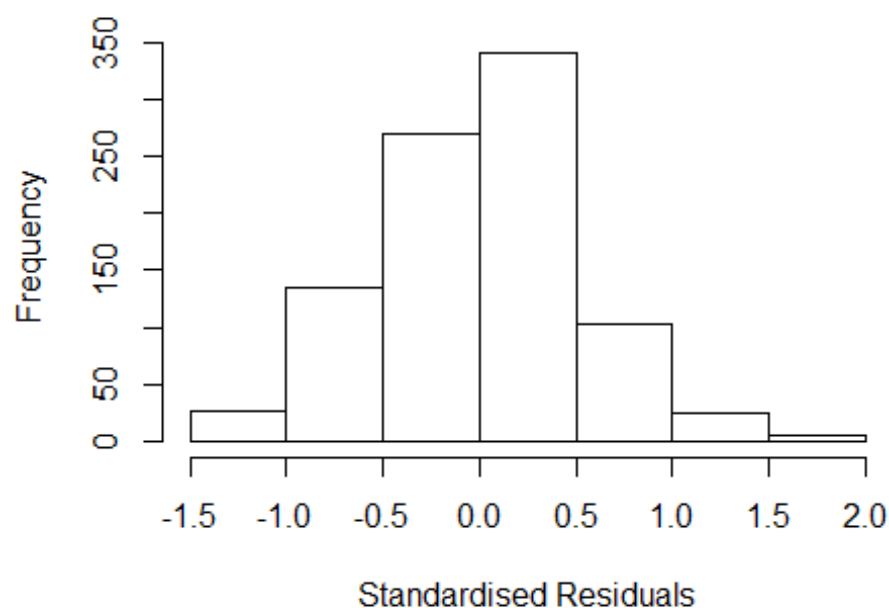


```

## Year2000          0.0403      0.1312      0.31      0.759
## Year2001         -0.0684      0.1460     -0.47      0.640
## Year2002         -0.2105      0.1286     -1.64      0.102
## Year2003         -0.2475      0.1434     -1.73      0.085 .
## Year2004          0.2544      0.1332      1.91      0.056 .
## Year2005          0.1129      0.1330      0.85      0.396
## Year2006          0.1168      0.1262      0.93      0.355
## Year2007          0.1004      0.1227      0.82      0.414
## Year2008          0.1336      0.1237      1.08      0.280
## Year2009          0.1082      0.1225      0.88      0.377
## Year2010          0.0760      0.1209      0.63      0.530
## Year2011          0.0520      0.1278      0.41      0.684
## Year2012          0.0954      0.1191      0.80      0.423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.274, Adjusted R-squared:  0.256
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 841 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.105  0.847  0.943  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      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 1.183 1      1.088
## LastAuthorFemale  1.099 1      1.048
## Year              1.250 16      1.007

```

Residuals from first and last author



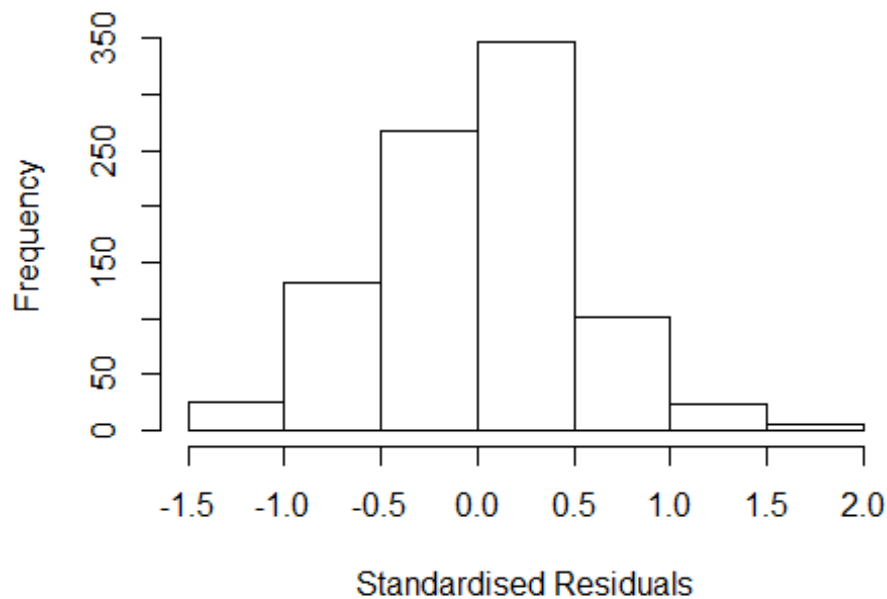
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4520 -0.3616 0.0347 0.3447 1.7001
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9260 0.1518 6.10 1.6e-09 ***
## FirstAuthorFemale1 0.0591 0.0420 1.41 0.1597
## LastAuthorFemale1 -0.0376 0.0577 -0.65 0.5149
## Year1997 0.3394 0.2181 1.56 0.1199
## Year1998 0.3714 0.1739 2.13 0.0330 *
## Year1999 0.3487 0.1708 2.04 0.0414 *
## Year2000 0.2580 0.1657 1.56 0.1199
## Year2001 0.1710 0.1722 0.99 0.3209
## Year2002 -0.2496 0.1873 -1.33 0.1829
## Year2003 -0.1671 0.2100 -0.80 0.4265
## Year2004 0.5045 0.1682 3.00 0.0028 **
## Year2005 0.3419 0.1688 2.03 0.0432 *
```

```

## Year2006          0.3375      0.1676      2.01      0.0443 *
## Year2007          0.3245      0.1617      2.01      0.0451 *
## Year2008          0.3885      0.1622      2.39      0.0168 *
## Year2009          0.3517      0.1619      2.17      0.0301 *
## Year2010          0.3046      0.1630      1.87      0.0620 .
## Year2011          0.3162      0.1614      1.96      0.0504 .
## Year2012          0.3695      0.1592      2.32      0.0205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.142, Adjusted R-squared:  0.124
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 829 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.156  0.839   0.940   0.884   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      1.11e-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 1.155 1      1.075
## Year              1.155 16      1.005

```

Residuals from first author

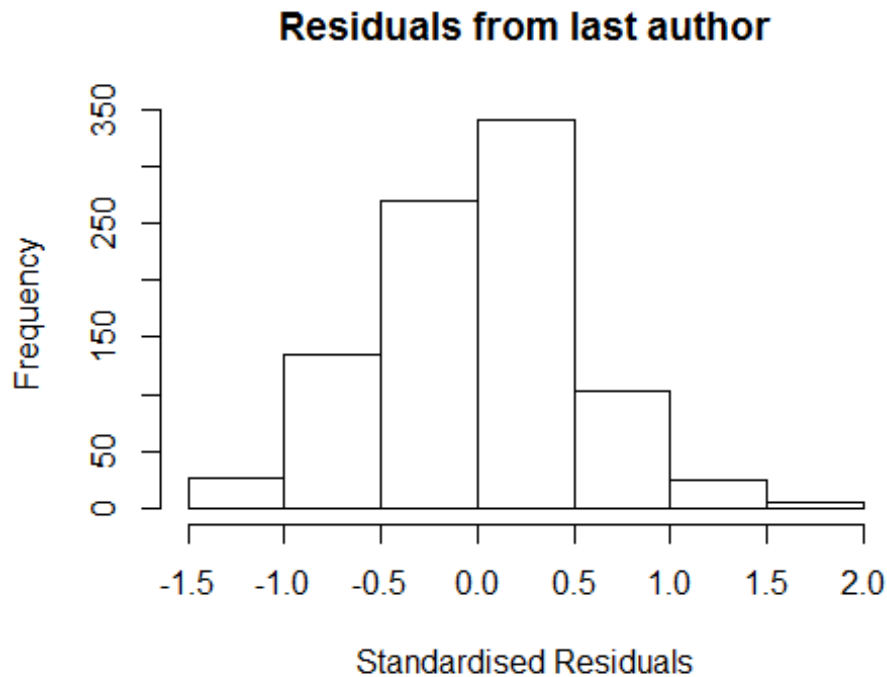


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4766 -0.3619 0.0354 0.3406 1.7024
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9263 0.1504 6.16 1.1e-09 ***
## FirstAuthorFemale1 0.0533 0.0417 1.28 0.2012
## Year1997 0.3396 0.2132 1.59 0.1115
## Year1998 0.3711 0.1715 2.16 0.0307 *
## Year1999 0.3479 0.1688 2.06 0.0396 *
## Year2000 0.2557 0.1647 1.55 0.1209
## Year2001 0.1704 0.1711 1.00 0.3196
## Year2002 -0.2577 0.1890 -1.36 0.1730
## Year2003 -0.1697 0.2104 -0.81 0.4202
## Year2004 0.4971 0.1653 3.01 0.0027 **
## Year2005 0.3380 0.1663 2.03 0.0423 *
## Year2006 0.3319 0.1649 2.01 0.0445 *
```

```

## Year2007          0.3209      0.1588      2.02      0.0436 *
## Year2008          0.3849      0.1602      2.40      0.0165 *
## Year2009          0.3467      0.1593      2.18      0.0298 *
## Year2010          0.3009      0.1605      1.87      0.0612 .
## Year2011          0.3122      0.1599      1.95      0.0512 .
## Year2012          0.3665      0.1580      2.32      0.0206 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.141, Adjusted R-squared:  0.125
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 833 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.159  0.840   0.940   0.886   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      1.11e-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 1.082 1          1.040
## Year              1.082 16          1.002

```



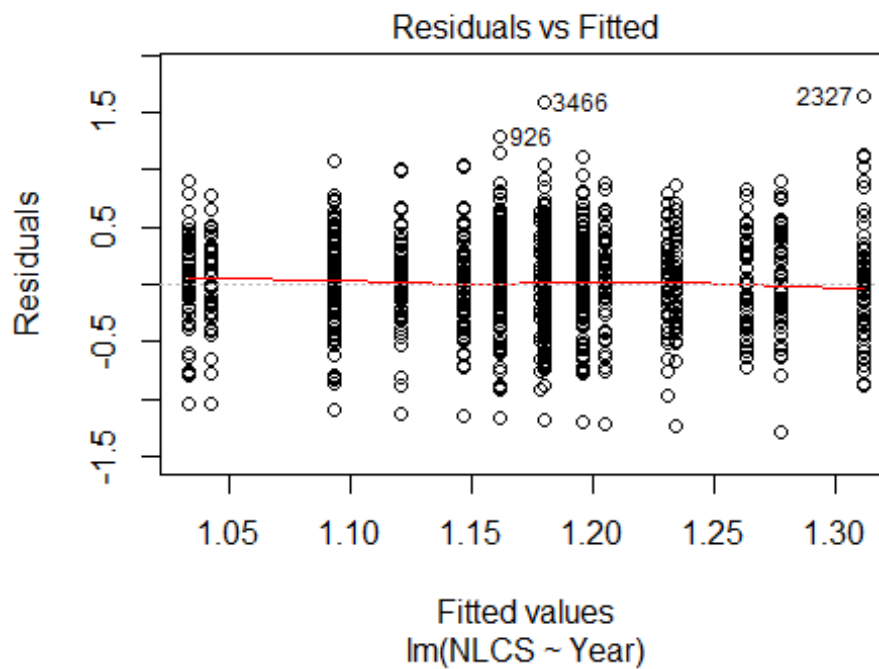
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4253 -0.3634 0.0328 0.3456 1.6880
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9285 0.1525 6.09 1.7e-09 ***
## LastAuthorFemale1 -0.0216 0.0567 -0.38 0.703
## Year1997 0.3541 0.2157 1.64 0.101
## Year1998 0.3798 0.1729 2.20 0.028 *
## Year1999 0.3532 0.1707 2.07 0.039 *
## Year2000 0.2723 0.1663 1.64 0.102
## Year2001 0.1762 0.1730 1.02 0.309
## Year2002 -0.2455 0.1920 -1.28 0.201
## Year2003 -0.1576 0.2134 -0.74 0.460
## Year2004 0.5184 0.1673 3.10 0.002 **
## Year2005 0.3485 0.1685 2.07 0.039 *
## Year2006 0.3621 0.1665 2.17 0.030 *
```

```

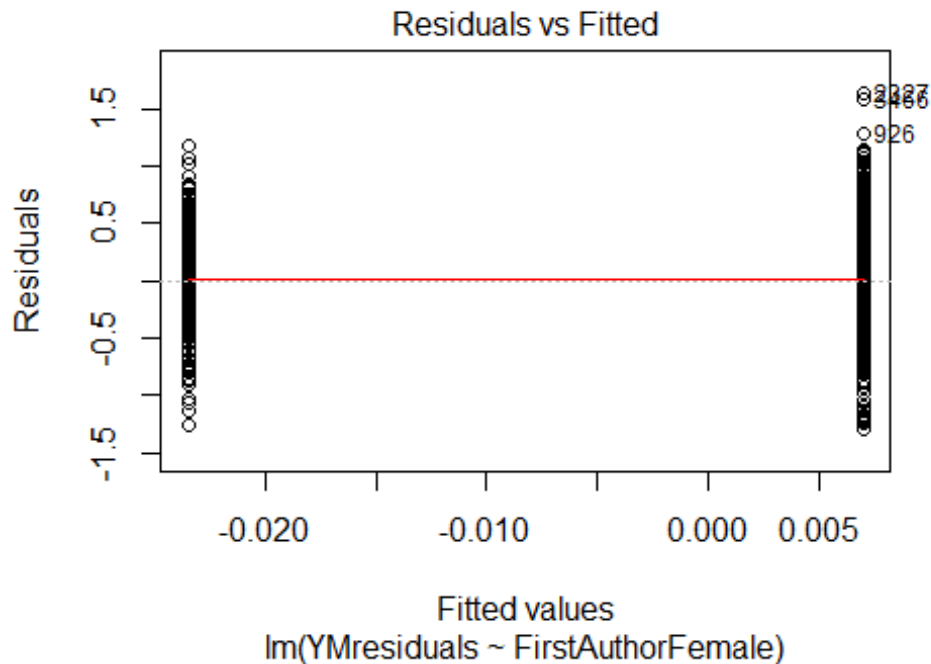
## Year2007          0.3433      0.1611      2.13      0.033 *
## Year2008          0.3987      0.1622      2.46      0.014 *
## Year2009          0.3626      0.1613      2.25      0.025 *
## Year2010          0.3179      0.1626      1.95      0.051 .
## Year2011          0.3226      0.1624      1.99      0.047 *
## Year2012          0.3771      0.1603      2.35      0.019 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.141, Adjusted R-squared:  0.125
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 835 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.162  0.840   0.941   0.885   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      1.11e-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: 902"
## [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
## 195 143 133 120 151 141 120 105 127 126 177 175 181 218 241
## 2011 2012
## 264 236
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62 58 50 51 58 50 64 32 46 49 61 75 77 102 117
## 2011 2012

```

```
## 131 124
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 51 43 49 55 39 56 30 41 42 55 64 71 78 98
## 2011 2012
## 115 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
```

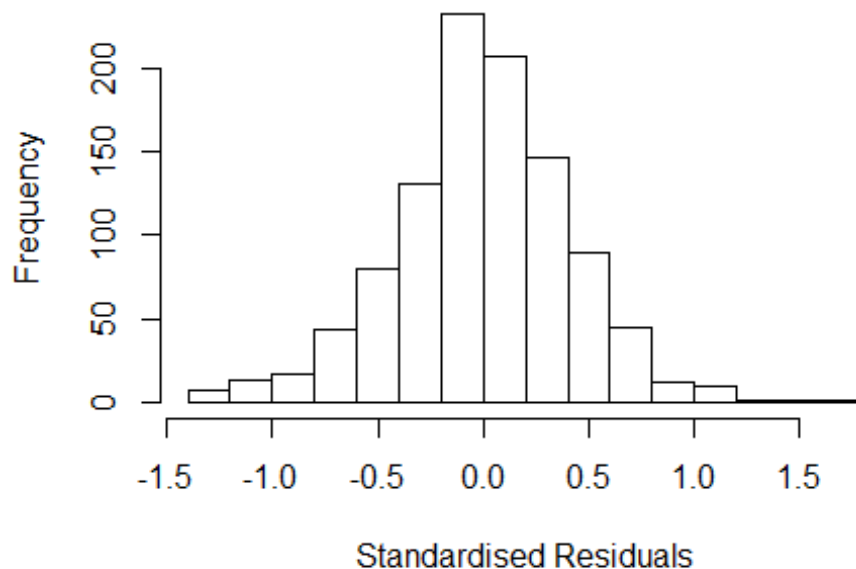


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.17, df = 1, p-value = 0.7
```

```
## [1] "Female first author team size 2018 geometric mean: 4.38343491053969"
## [1] "Male first author team size 2018 geometric mean: 4.07706044563314"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 620, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.49328664536681"
## [1] "Male last author team size 2018 geometric mean: 4.0437627261218"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 720, 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 1.083 1      1.041
## LastAuthorFemale  1.092 1      1.045
## UniqueAuthors    1.522 4      1.054
## Year             1.694 16      1.017
```

Residuals from first and last author and team size



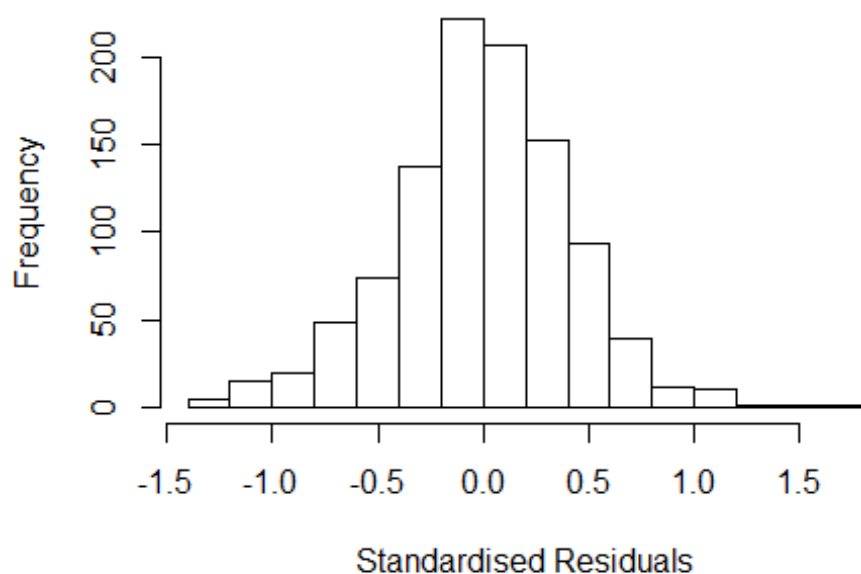
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29431 -0.24176 -0.00273 0.25120 1.73303
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22205 0.09053 13.50 <2e-16 ***
## FirstAuthorFemale1 -0.05006 0.03167 -1.58 0.114
## LastAuthorFemale1 -0.03892 0.04092 -0.95 0.342
## UniqueAuthors2 0.01790 0.06957 0.26 0.797
## UniqueAuthors3 0.07606 0.06827 1.11 0.265
## UniqueAuthors4 0.10279 0.06929 1.48 0.138
## UniqueAuthors5 0.12233 0.06991 1.75 0.080 .
## Year1997 0.00552 0.08348 0.07 0.947
## Year1998 -0.01949 0.08017 -0.24 0.808
## Year1999 -0.03660 0.08214 -0.45 0.656
```

```

## Year2000      -0.18817    0.09222   -2.04    0.042 *
## Year2001      -0.06933    0.12537   -0.55    0.580
## Year2002      -0.03897    0.08865   -0.44    0.660
## Year2003      -0.06147    0.09617   -0.64    0.523
## Year2004      -0.17355    0.08176   -2.12    0.034 *
## Year2005      -0.02488    0.08355   -0.30    0.766
## Year2006      -0.15476    0.07888   -1.96    0.050 .
## Year2007      -0.10623    0.07799   -1.36    0.173
## Year2008      -0.02398    0.08099   -0.30    0.767
## Year2009      -0.12845    0.07679   -1.67    0.095 .
## Year2010      -0.09658    0.07170   -1.35    0.178
## Year2011      -0.10598    0.06681   -1.59    0.113
## Year2012      -0.09979    0.07190   -1.39    0.165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0334, Adjusted R-squared:  0.0124
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 937 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0022 0.8530 0.9500 0.8900 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          9.62e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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 1.075 1 1.037
## Year 1.143 16 1.004

```

Residuals from first and last author



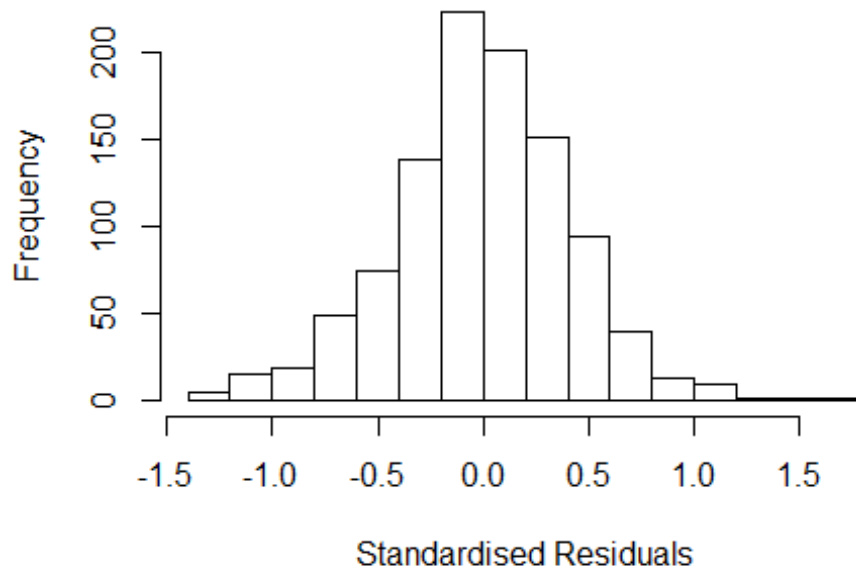
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3028 -0.2487 -0.0055  0.2560  1.6784
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2922     0.0598   21.61  <2e-16 ***
## FirstAuthorFemale1 -0.0432     0.0316   -1.37    0.172
## LastAuthorFemale1 -0.0325     0.0405   -0.80    0.422
## Year1997          0.0106     0.0829    0.13    0.898
## Year1998         -0.0346     0.0799   -0.43    0.665
## Year1999         -0.0454     0.0823   -0.55    0.581
## Year2000         -0.2121     0.0924   -2.30    0.022 *
## Year2001         -0.1011     0.1220   -0.83    0.408
## Year2002         -0.0443     0.0875   -0.51    0.613
## Year2003         -0.0502     0.0937   -0.54    0.592
## Year2004         -0.1671     0.0815   -2.05    0.041 *
## Year2005         -0.0144     0.0830   -0.17    0.862
```

```

## Year2006          -0.1570      0.0770   -2.04    0.042 *
## Year2007          -0.1061      0.0789   -1.35    0.179
## Year2008          -0.0216      0.0813   -0.27    0.790
## Year2009          -0.1184      0.0759   -1.56    0.119
## Year2010          -0.0823      0.0712   -1.16    0.248
## Year2011          -0.0947      0.0669   -1.42    0.157
## Year2012          -0.0912      0.0722   -1.26    0.207
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.0232, Adjusted R-squared:  0.00592
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 950 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0145 0.8570 0.9540 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      9.62e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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      1.032
## Year              1.066 16      1.002

```

Residuals from first author



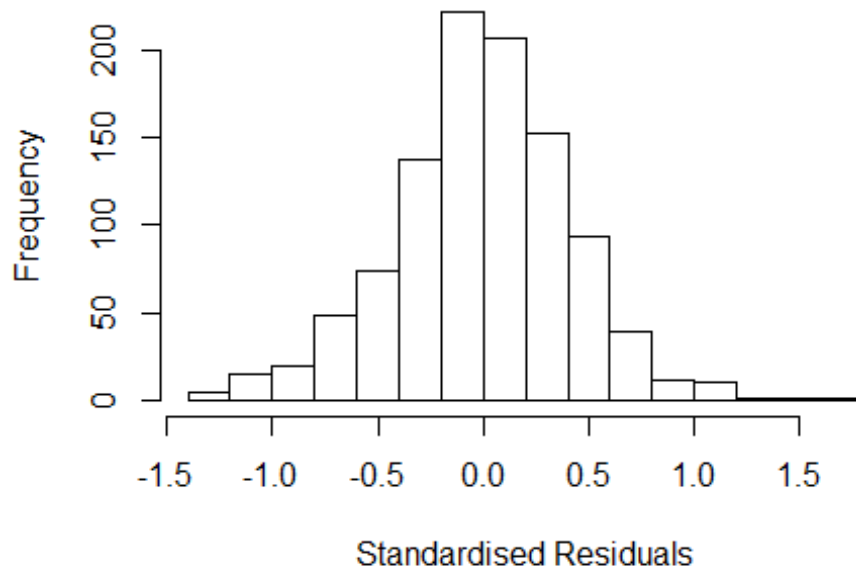
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3020 -0.2458 -0.0049 0.2585 1.6814
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2908 0.0600 21.53 <2e-16 ***
## FirstAuthorFemale1 -0.0443 0.0316 -1.40 0.161
## Year1997 0.0112 0.0831 0.13 0.893
## Year1998 -0.0361 0.0802 -0.45 0.652
## Year1999 -0.0462 0.0825 -0.56 0.576
## Year2000 -0.2128 0.0925 -2.30 0.022 *
## Year2001 -0.1008 0.1227 -0.82 0.411
## Year2002 -0.0472 0.0877 -0.54 0.590
## Year2003 -0.0546 0.0942 -0.58 0.562
## Year2004 -0.1668 0.0819 -2.04 0.042 *
## Year2005 -0.0199 0.0829 -0.24 0.810
## Year2006 -0.1600 0.0771 -2.07 0.038 *
```

```

## Year2007          -0.1105      0.0787   -1.41    0.160
## Year2008          -0.0232      0.0813   -0.29    0.775
## Year2009          -0.1198      0.0763   -1.57    0.117
## Year2010          -0.0849      0.0712   -1.19    0.233
## Year2011          -0.0969      0.0671   -1.45    0.149
## Year2012          -0.0943      0.0724   -1.30    0.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.00608
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 950 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0133 0.8540 0.9540 0.8920 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      9.62e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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

```

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
## -1.297592 -0.251278 0.000226 0.256935 1.688373
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2866 0.0595 21.61 <2e-16 ***
## LastAuthorFemale1 -0.0353 0.0406 -0.87 0.385
## Year1997 0.0110 0.0835 0.13 0.896
## Year1998 -0.0327 0.0795 -0.41 0.681
## Year1999 -0.0463 0.0826 -0.56 0.575
## Year2000 -0.2145 0.0917 -2.34 0.020 *
## Year2001 -0.1012 0.1221 -0.83 0.407
## Year2002 -0.0474 0.0873 -0.54 0.587
## Year2003 -0.0568 0.0954 -0.60 0.551
## Year2004 -0.1669 0.0816 -2.05 0.041 *
## Year2005 -0.0170 0.0830 -0.20 0.838
## Year2006 -0.1666 0.0769 -2.17 0.031 *
```

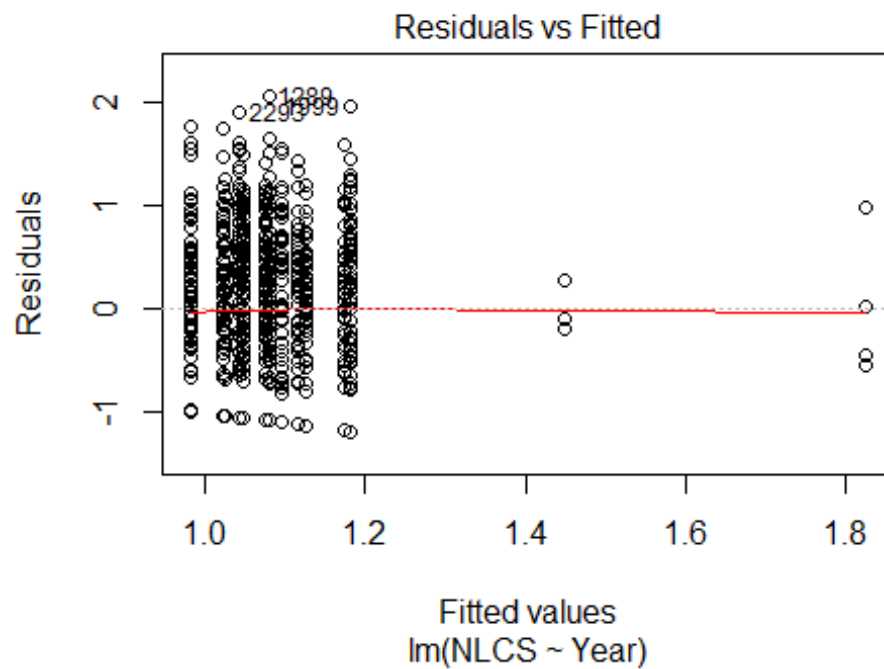


```

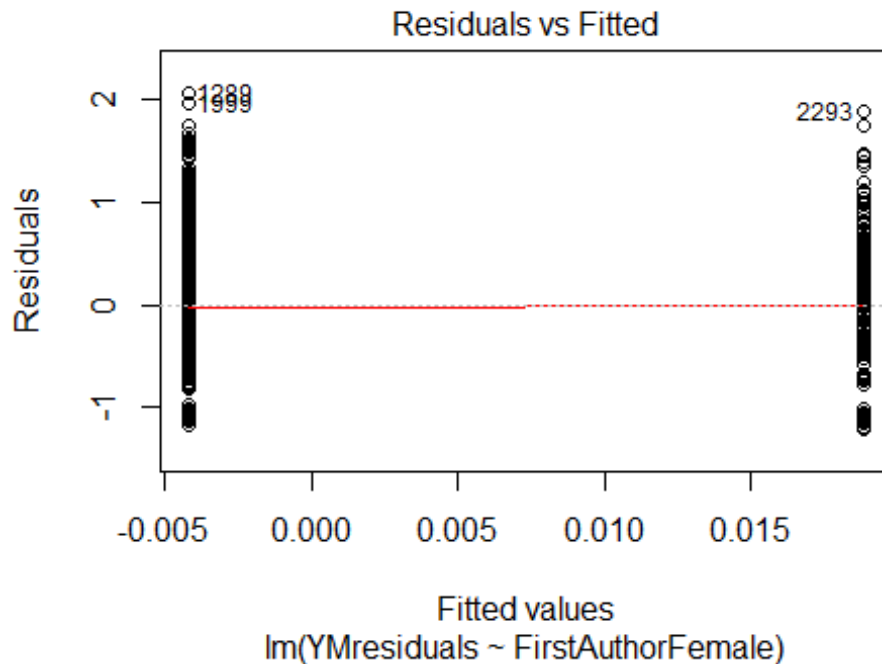
## Year2007          -0.1135      0.0782    -1.45      0.147
## Year2008          -0.0260      0.0813    -0.32      0.749
## Year2009          -0.1261      0.0757    -1.67      0.096 .
## Year2010          -0.0874      0.0711    -1.23      0.220
## Year2011          -0.1003      0.0670    -1.50      0.135
## Year2012          -0.0964      0.0724    -1.33      0.183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.0211, Adjusted R-squared:  0.00477
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 948 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0143 0.8570 0.9530 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      9.62e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1039"
## [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
## 108 61 10 131 11 71 79 91 100 92 127 122 135 149 133
## 2011 2012
## 143 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 19 3 55 4 30 40 44 57 47 61 75 86 98 96
## 2011 2012

```

```
##      89      87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   16    3   49    3   27   32   40   49   42   58   72   77   86   88
## 2011 2012
##   81   75
## [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
```

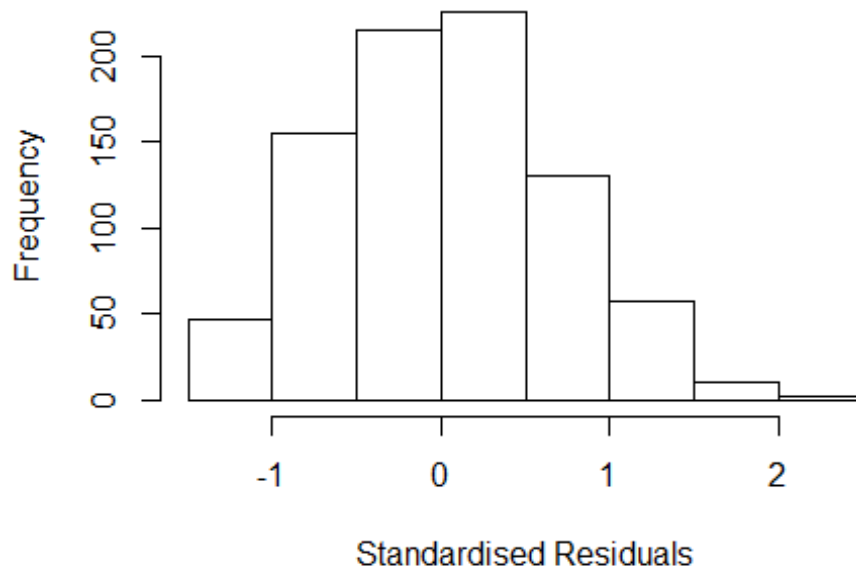


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.23, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 1.94328331572615"
## [1] "Male first author team size 2018 geometric mean: 1.4653415838455"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 930, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.63089529958951"
## [1] "Male last author team size 2018 geometric mean: 1.54644341046875"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 620, 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 1.637 1      1.279
## LastAuthorFemale  1.625 1      1.275
## UniqueAuthors    1.629 4      1.063
## Year              1.697 16     1.017
```

Residuals from first and last author and team size



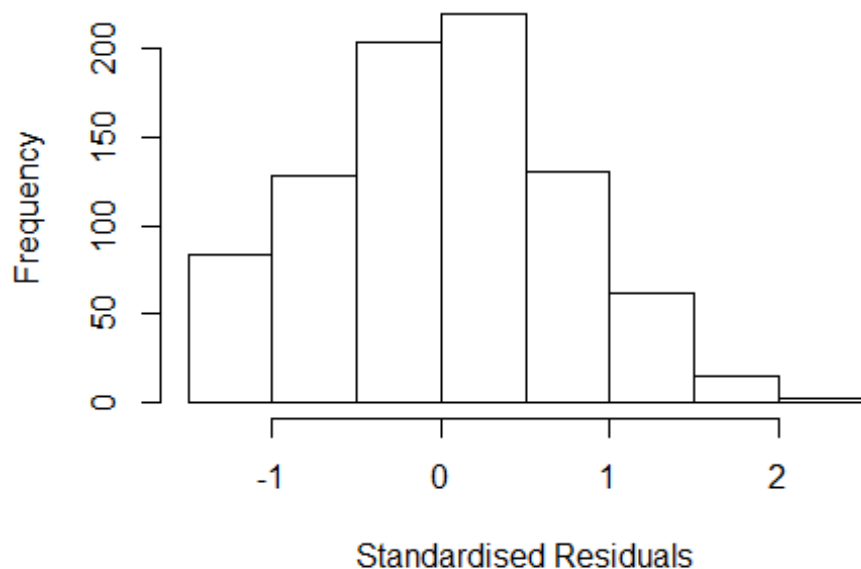
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3739 -0.4906 0.0157 0.4633 2.2048
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03339 0.10935 9.45 < 2e-16 ***
## FirstAuthorFemale1 -0.03306 0.07691 -0.43 0.6674
## LastAuthorFemale1 -0.04878 0.07933 -0.61 0.5387
## UniqueAuthors2 0.34209 0.05938 5.76 1.2e-08 ***
## UniqueAuthors3 0.27216 0.08743 3.11 0.0019 **
## UniqueAuthors4 0.24649 0.19201 1.28 0.1996
## UniqueAuthors5 0.41874 0.17119 2.45 0.0147 *
## Year1997 -0.05751 0.19470 -0.30 0.7678
## Year1998 0.41484 0.15870 2.61 0.0091 **
## Year1999 -0.15600 0.14912 -1.05 0.2958
```

```

## Year2000      0.95670      0.39296      2.43      0.0151 *
## Year2001     -0.08459      0.16191     -0.52      0.6015
## Year2002     -0.23615      0.15389     -1.53      0.1253
## Year2003     -0.14031      0.14771     -0.95      0.3424
## Year2004      0.02452      0.14798      0.17      0.8684
## Year2005     -0.26911      0.14709     -1.83      0.0677 .
## Year2006     -0.01870      0.13350     -0.14      0.8886
## Year2007     -0.08270      0.13976     -0.59      0.5542
## Year2008     -0.11451      0.13304     -0.86      0.3896
## Year2009     -0.09619      0.13072     -0.74      0.4620
## Year2010     -0.15452      0.14137     -1.09      0.2747
## Year2011     -0.00161      0.14312     -0.01      0.9910
## Year2012     -0.20134      0.14923     -1.35      0.1776
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.728
## Multiple R-squared:  0.0644, Adjusted R-squared:  0.0394
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 782 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.338  0.874  0.953   0.918   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      1.18e-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  1.582 1      1.258
## Year              1.171 16      1.005

```

Residuals from first and last author



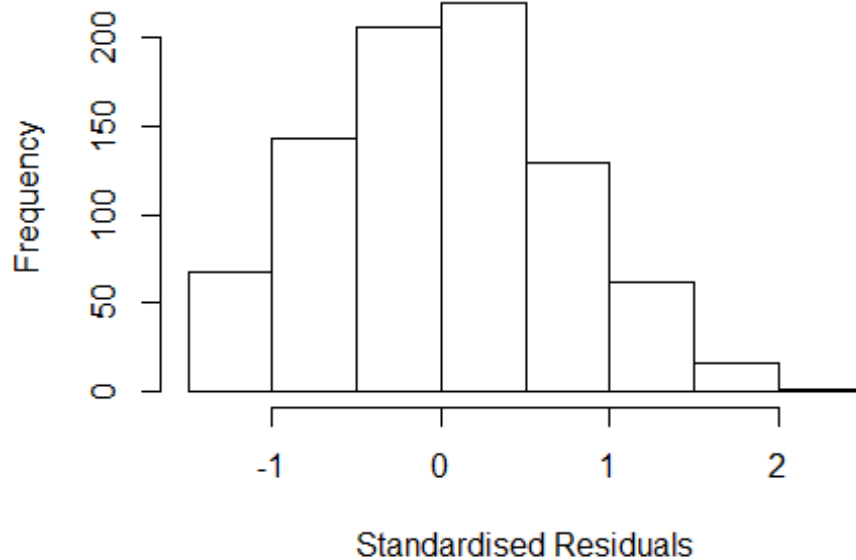
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16781 -0.50035 0.00835 0.48917 2.08150
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16082 0.10679 10.87 <2e-16 ***
## FirstAuthorFemale1 0.00224 0.07841 0.03 0.977
## LastAuthorFemale1 -0.03118 0.08112 -0.38 0.701
## Year1997 -0.07953 0.19224 -0.41 0.679
## Year1998 0.28748 0.15688 1.83 0.067 .
## Year1999 -0.16406 0.15564 -1.05 0.292
## Year2000 0.83091 0.38905 2.14 0.033 *
## Year2001 -0.11456 0.17056 -0.67 0.502
## Year2002 -0.27248 0.15159 -1.80 0.073 .
## Year2003 -0.14813 0.14794 -1.00 0.317
## Year2004 -0.07285 0.14781 -0.49 0.622
## Year2005 -0.26566 0.14559 -1.82 0.068 .
```

```

## Year2006      -0.03286    0.13407   -0.25    0.806
## Year2007      -0.10132    0.13715   -0.74    0.460
## Year2008      -0.10544    0.13522   -0.78    0.436
## Year2009      -0.06626    0.13317   -0.50    0.619
## Year2010      -0.15844    0.14137   -1.12    0.263
## Year2011       0.00699    0.14227    0.05    0.961
## Year2012      -0.18116    0.15061   -1.20    0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.762
## Multiple R-squared:  0.0182, Adjusted R-squared:  -0.00322
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.436  0.855   0.954   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      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 1.123 1      1.060
## Year              1.123 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
## -1.1636 -0.4979 0.0113 0.4927 2.0831
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15928 0.10737 10.80 <2e-16 ***
## FirstAuthorFemale1 -0.01579 0.06583 -0.24 0.810
## Year1997 -0.07896 0.19245 -0.41 0.682
## Year1998 0.28901 0.15727 1.84 0.066 .
## Year1999 -0.16434 0.15594 -1.05 0.292
## Year2000 0.83244 0.38924 2.14 0.033 *
## Year2001 -0.11396 0.17100 -0.67 0.505
## Year2002 -0.27221 0.15161 -1.80 0.073 .
## Year2003 -0.14895 0.14801 -1.01 0.315
## Year2004 -0.07315 0.14823 -0.49 0.622
## Year2005 -0.26465 0.14645 -1.81 0.071 .
## Year2006 -0.03354 0.13443 -0.25 0.803
```

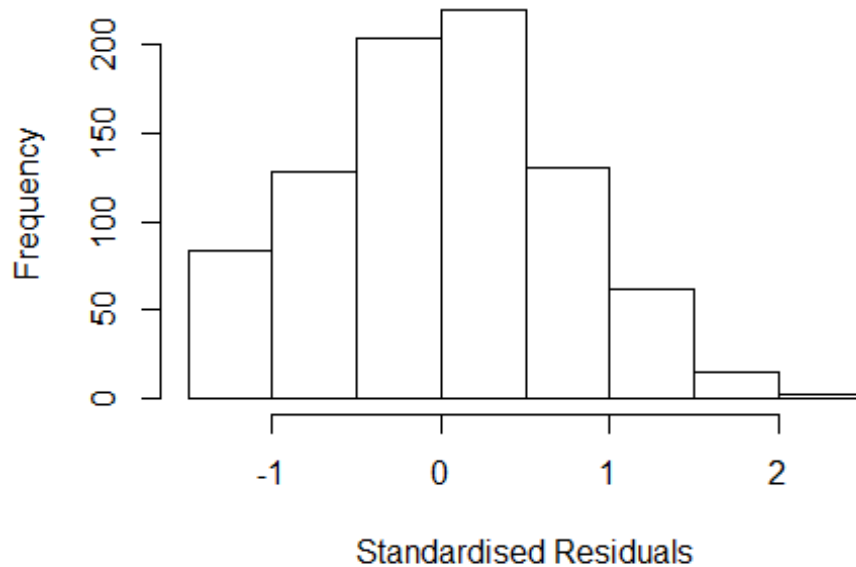


```

## Year2007      -0.10142    0.13770   -0.74    0.462
## Year2008      -0.10372    0.13543   -0.77    0.444
## Year2009      -0.06814    0.13324   -0.51    0.609
## Year2010      -0.15955    0.14164   -1.13    0.260
## Year2011       0.00434    0.14235    0.03    0.976
## Year2012      -0.18399    0.15045   -1.22    0.222
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.762
## Multiple R-squared:  0.0181, Adjusted R-squared:  -0.00215
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 773 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.435  0.856  0.954  0.921  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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 1.106 1      1.052
## Year      1.106 16      1.003

```

Residuals from last author



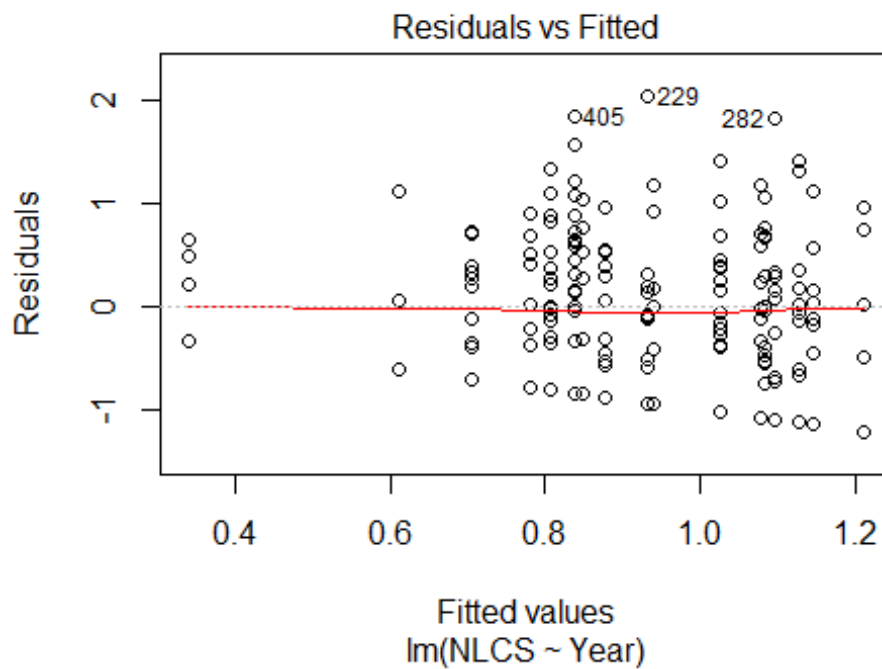
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16794 -0.50048 0.00839 0.48909 2.08143
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16100 0.10692 10.86 <2e-16 ***
## LastAuthorFemale1 -0.02983 0.06781 -0.44 0.660
## Year1997 -0.07968 0.19238 -0.41 0.679
## Year1998 0.28729 0.15697 1.83 0.068 .
## Year1999 -0.16422 0.15568 -1.05 0.292
## Year2000 0.83069 0.38917 2.13 0.033 *
## Year2001 -0.11473 0.17065 -0.67 0.502
## Year2002 -0.27250 0.15162 -1.80 0.073 .
## Year2003 -0.14800 0.14771 -1.00 0.317
## Year2004 -0.07293 0.14792 -0.49 0.622
## Year2005 -0.26579 0.14565 -1.82 0.068 .
## Year2006 -0.03290 0.13420 -0.25 0.806
```

```

## Year2007          -0.10143      0.13726    -0.74      0.460
## Year2008          -0.10533      0.13496    -0.78      0.435
## Year2009          -0.06613      0.13280    -0.50      0.619
## Year2010          -0.15844      0.14138    -1.12      0.263
## Year2011           0.00694      0.14241      0.05      0.961
## Year2012          -0.18139      0.15066    -1.20      0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.761
## Multiple R-squared:  0.0182, Adjusted R-squared:  -0.00199
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.435  0.855  0.953  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      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: 844"
## [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
##   14   9  11  12  17  17  18  18  25  20  22  41  33  30  16
## 2011 2012
##   59  63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    5    8    9   13    7   10   12   13    5   11   12   10   18   11
## 2011 2012

```

```
## 29 28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 5 8 8 12 7 9 11 11 4 10 9 10 15 10
## 2011 2012
## 25 27
## [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
```



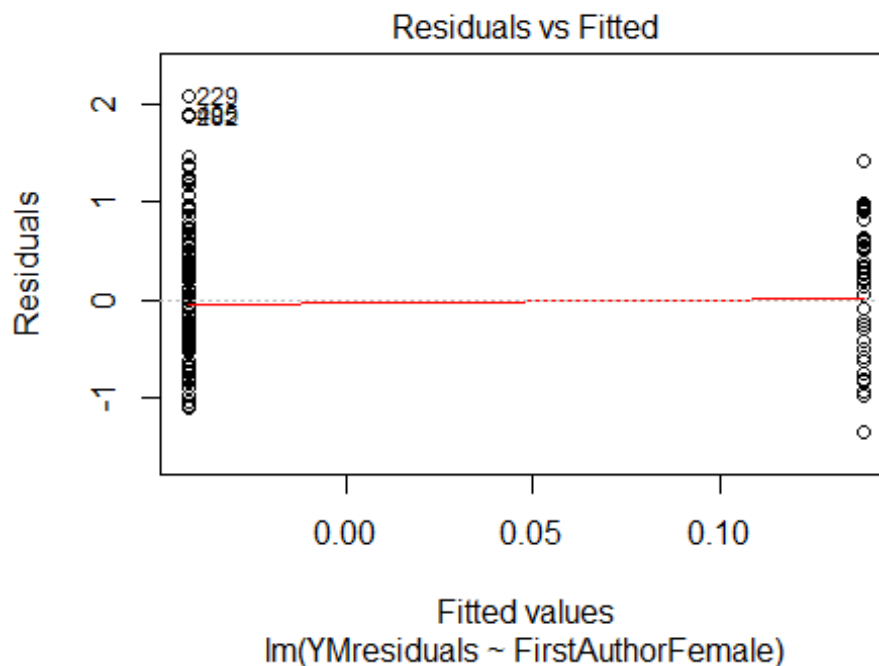
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.24, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 2.1689435423954"
## [1] "Male first author team size 2018 geometric mean: 1.54590939849101"

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

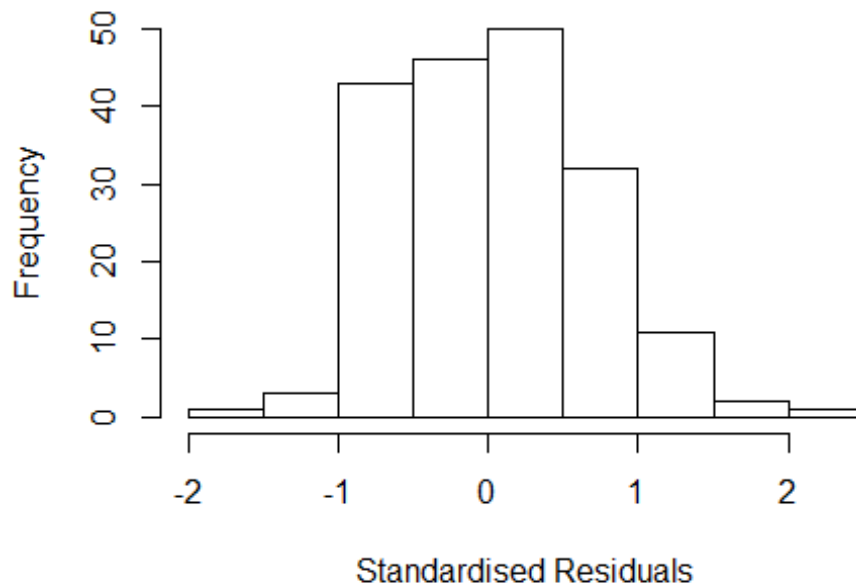
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 40, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2"
## [1] "Male last author team size 2018 geometric mean: 1.64375182951723"

## 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] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.708  1      1.646
## LastAuthorFemale  2.823  1      1.680
## UniqueAuthors    3.543  4      1.171
## Year              5.139 16      1.052
```

Residuals from first and last author and team size



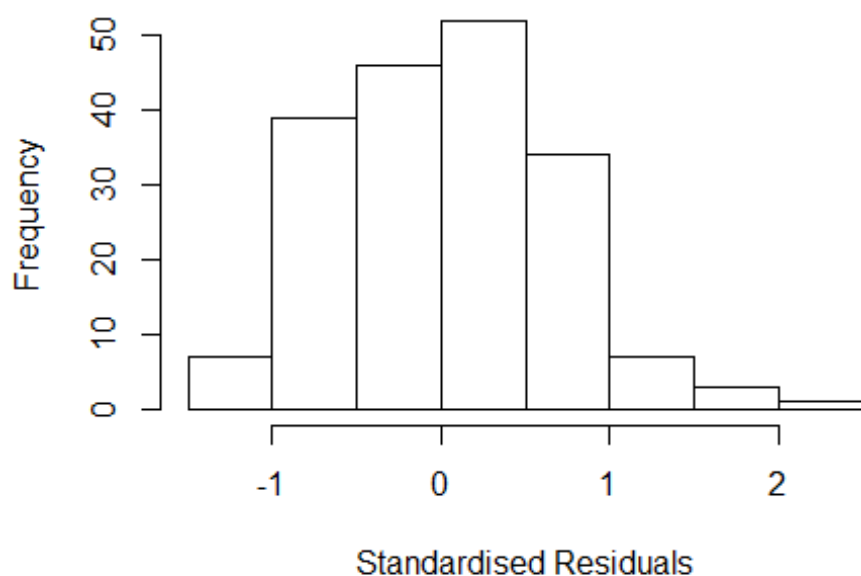
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5078 -0.4896 0.0138 0.4938 2.2748
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8970 0.2872 3.12 0.0021 **
## FirstAuthorFemale1 0.1053 0.1963 0.54 0.5922
## LastAuthorFemale1 0.2055 0.2225 0.92 0.3571
## UniqueAuthors2 0.1058 0.1247 0.85 0.3974
## UniqueAuthors3 0.1790 0.1868 0.96 0.3394
## UniqueAuthors4 0.3463 0.2826 1.23 0.2222
## UniqueAuthors5 -0.0359 0.5421 -0.07 0.9473
## Year1997 0.1942 0.6273 0.31 0.7572
## Year1998 0.1632 0.3707 0.44 0.6603
## Year1999 -0.1066 0.3687 -0.29 0.7730
```

```

## Year2000          0.1209      0.3384      0.36      0.7213
## Year2001         -0.6930      0.3176     -2.18      0.0305 *
## Year2002         -0.2145      0.3936     -0.55      0.5864
## Year2003         -0.4188      0.3436     -1.22      0.2246
## Year2004         -0.2646      0.3458     -0.77      0.4452
## Year2005         -0.5572      0.5986     -0.93      0.3533
## Year2006          0.0793      0.4100      0.19      0.8469
## Year2007         -0.2018      0.3544     -0.57      0.5699
## Year2008          0.0731      0.3610      0.20      0.8398
## Year2009         -0.1745      0.3280     -0.53      0.5955
## Year2010          0.0917      0.3689      0.25      0.8041
## Year2011         -0.1918      0.3532     -0.54      0.5879
## Year2012         -0.2540      0.3196     -0.79      0.4279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.654
## Multiple R-squared:  0.113, Adjusted R-squared:  -0.00482
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.201  0.878  0.931  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      5.29e-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 2.269 1      1.506
## LastAuthorFemale  2.202 1      1.484
## Year              1.783 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
## -1.4265 -0.4854 0.0201 0.4713 2.2088
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9209 0.2877 3.20 0.0016 **
## FirstAuthorFemale1 0.1639 0.1836 0.89 0.3734
## LastAuthorFemale1 0.1485 0.1964 0.76 0.4506
## Year1997 0.1933 0.5833 0.33 0.7408
## Year1998 0.1785 0.3562 0.50 0.6170
## Year1999 -0.0908 0.3663 -0.25 0.8046
## Year2000 0.1523 0.3347 0.46 0.6495
## Year2001 -0.6722 0.3116 -2.16 0.0324 *
## Year2002 -0.1953 0.3792 -0.51 0.6073
## Year2003 -0.3566 0.3320 -1.07 0.2844
## Year2004 -0.2282 0.3380 -0.68 0.5006
## Year2005 -0.5562 0.5951 -0.93 0.3513
```

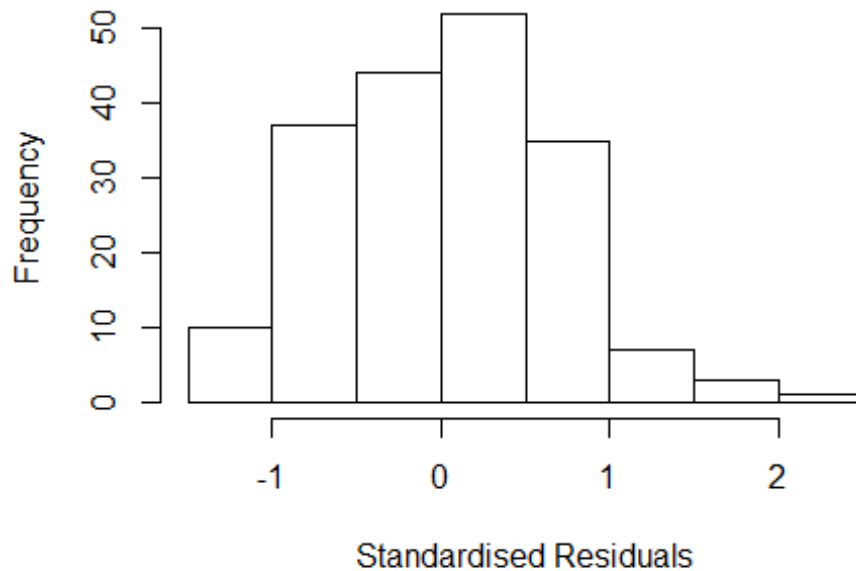


```

## Year2006          0.0967      0.4085      0.24      0.8132
## Year2007         -0.1597      0.3586     -0.45      0.6567
## Year2008          0.0967      0.3591      0.27      0.7881
## Year2009         -0.1421      0.3217     -0.44      0.6593
## Year2010          0.1353      0.3625      0.37      0.7094
## Year2011         -0.1445      0.3434     -0.42      0.6744
## Year2012         -0.1678      0.3097     -0.54      0.5887
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.721
## Multiple R-squared:  0.093, Adjusted R-squared:  -0.00306
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 168 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.327  0.897  0.942  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      5.29e-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 1.33 1      1.153
## Year              1.33 16      1.009

```

Residuals from first author



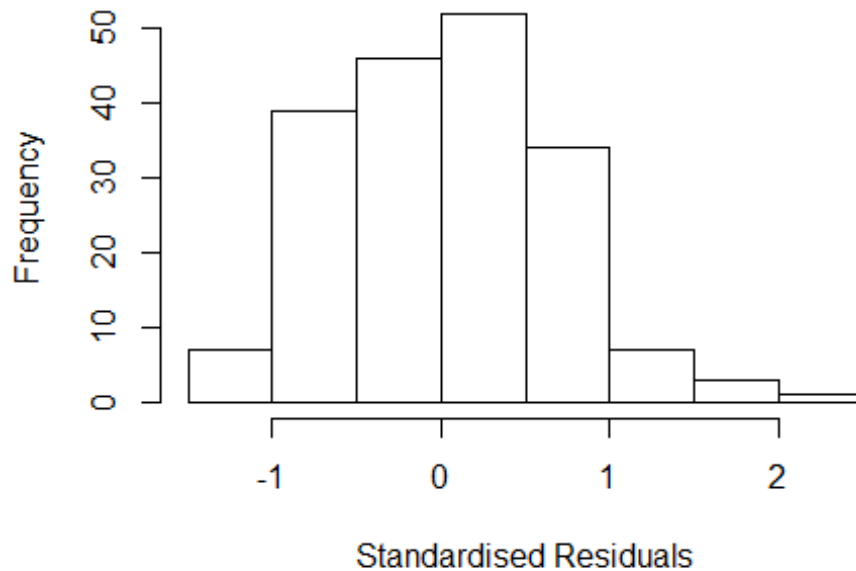
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3626 -0.4861 0.0212 0.4778 2.2270
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9198 0.2892 3.18 0.0017 **
## FirstAuthorFemale1 0.2411 0.1445 1.67 0.0970 .
## Year1997 0.2017 0.5514 0.37 0.7150
## Year1998 0.2084 0.3562 0.59 0.5593
## Year1999 -0.0728 0.3716 -0.20 0.8448
## Year2000 0.1450 0.3349 0.43 0.6655
## Year2001 -0.6515 0.3161 -2.06 0.0408 *
## Year2002 -0.2061 0.3793 -0.54 0.5876
## Year2003 -0.3524 0.3354 -1.05 0.2948
## Year2004 -0.1917 0.3322 -0.58 0.5647
## Year2005 -0.5180 0.5732 -0.90 0.3674
## Year2006 0.0804 0.4206 0.19 0.8486
```

```

## Year2007          -0.1768      0.3576   -0.49   0.6217
## Year2008           0.0944      0.3597    0.26   0.7934
## Year2009          -0.1077      0.3201   -0.34   0.7370
## Year2010           0.1449      0.3653    0.40   0.6922
## Year2011          -0.1529      0.3488   -0.44   0.6617
## Year2012          -0.1608      0.3114   -0.52   0.6062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.706
## Multiple R-squared:  0.0902, Adjusted R-squared:  -0.000225
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 168 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.299  0.889   0.945   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      5.29e-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 1.243 1          1.115
## Year            1.243 16          1.007

```

Residuals from last author



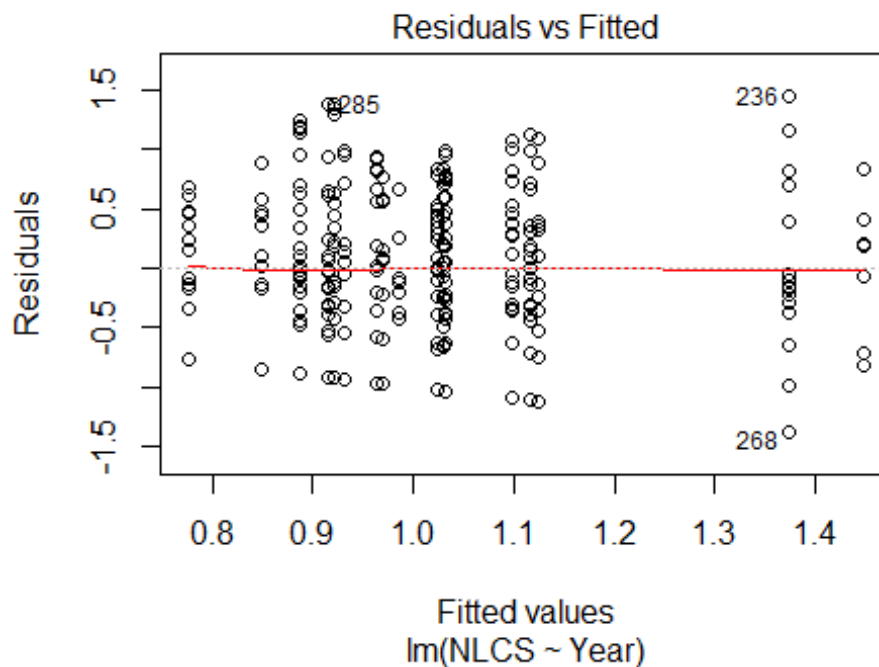
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4177 -0.4978 0.0119 0.4886 2.1910
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9197 0.2894 3.18 0.0018 **
## LastAuthorFemale1 0.2542 0.1533 1.66 0.0991 .
## Year1997 0.2438 0.6036 0.40 0.6868
## Year1998 0.1700 0.3610 0.47 0.6384
## Year1999 -0.0442 0.3747 -0.12 0.9063
## Year2000 0.1831 0.3430 0.53 0.5941
## Year2001 -0.6550 0.3159 -2.07 0.0396 *
## Year2002 -0.1462 0.3731 -0.39 0.6957
## Year2003 -0.3083 0.3295 -0.94 0.3508
## Year2004 -0.2428 0.3390 -0.72 0.4749
## Year2005 -0.5261 0.5799 -0.91 0.3656
## Year2006 0.1272 0.3910 0.33 0.7453
```

```

## Year2007          -0.1407      0.3546   -0.40    0.6921
## Year2008           0.0941      0.3598    0.26    0.7941
## Year2009          -0.1406      0.3252   -0.43    0.6659
## Year2010           0.1436      0.3654    0.39    0.6949
## Year2011          -0.1219      0.3497   -0.35    0.7279
## Year2012          -0.1526      0.3106   -0.49    0.6239
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.704
## Multiple R-squared:  0.0884, Adjusted R-squared:  -0.00225
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 165 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.313  0.887  0.939  0.913  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      5.29e-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: 189"
## [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
##   30   33   35   36   38   32   40   25   35   34   36   29   31   26   36
## 2011 2012
##   30   60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   11   13   12   20   20   19    9   15   22   25   25   15   16   26
## 2011 2012

```

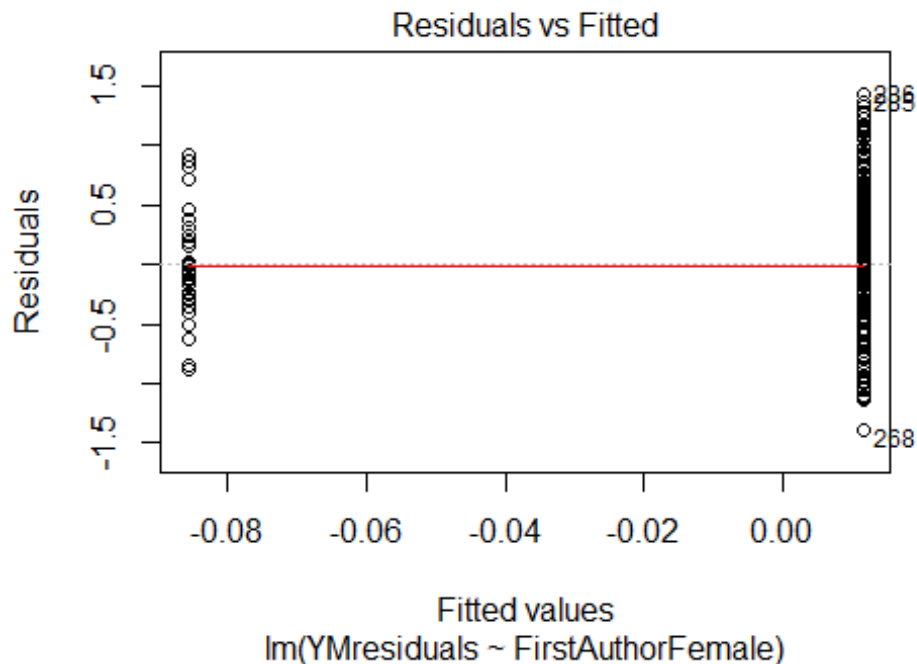
```
## 23 40
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 11 13 12 17 19 19 7 15 22 24 23 12 14 23
## 2011 2012
## 22 35
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: 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 = 7.8, df = 1, p-value = 0.005
## [1] "Female first author team size 2018 geometric mean: 2.82842712474619"
## [1] "Male first author team size 2018 geometric mean: 1.38741870573731"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

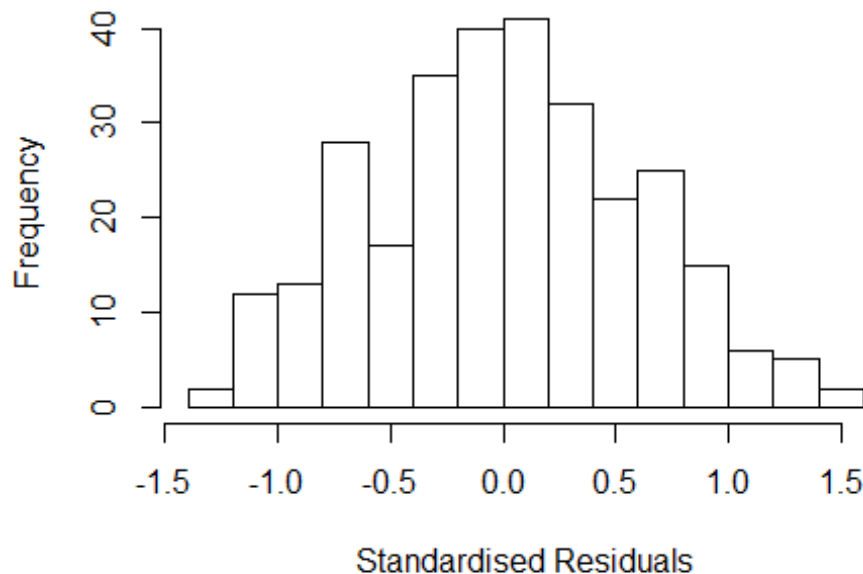
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 47, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.28942848510666"
## [1] "Male last author team size 2018 geometric mean: 1.38326161560748"

## 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 = 63, 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  3.198  1      1.788
## LastAuthorFemale 11.516  1      3.394
## UniqueAuthors    46.899  4      1.618
## Year              31.562 16      1.114
```

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
## -1.295711 -0.391822 0.000996 0.406440 1.443396
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37000 0.21236 6.45 5e-10 ***
## FirstAuthorFemale1 -0.12685 0.11699 -1.08 0.2792
## LastAuthorFemale1 -0.05550 0.12661 -0.44 0.6615
## UniqueAuthors2 0.26490 0.09277 2.86 0.0046 **
## UniqueAuthors3 -0.02224 0.16409 -0.14 0.8923
## UniqueAuthors4 0.11486 0.26011 0.44 0.6592
## UniqueAuthors5 -0.00495 0.17546 -0.03 0.9775
## Year1997 -0.33923 0.28845 -1.18 0.2406
## Year1998 -0.43431 0.29431 -1.48 0.1412
## Year1999 -0.27360 0.27348 -1.00 0.3180
```

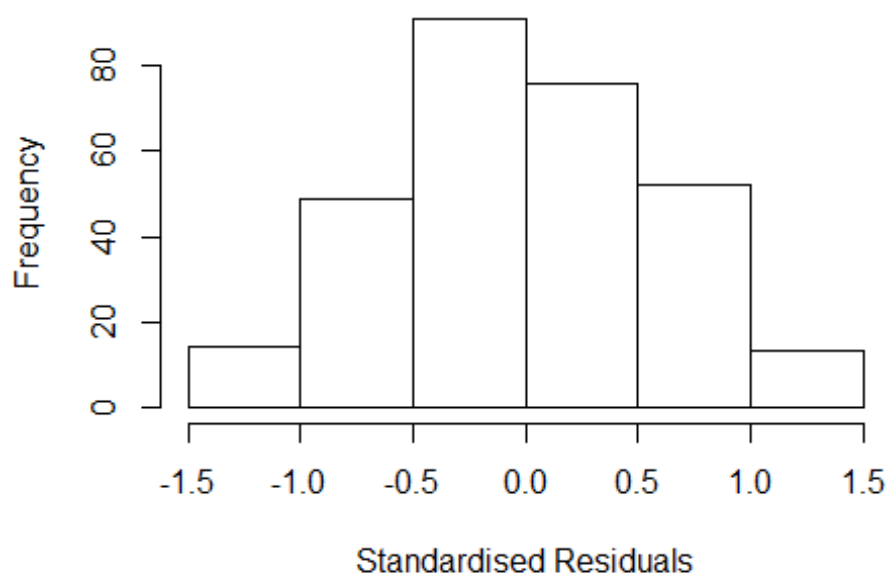


```

## Year2000      -0.42199      0.27812      -1.52      0.1303
## Year2001      -0.08922      0.26192      -0.34      0.7336
## Year2002      -0.55615      0.24666      -2.25      0.0249 *
## Year2003      -0.38149      0.25091      -1.52      0.1296
## Year2004      -0.37900      0.24225      -1.56      0.1189
## Year2005      -0.28156      0.24636      -1.14      0.2541
## Year2006      -0.34805      0.24105      -1.44      0.1499
## Year2007      -0.34435      0.24770      -1.39      0.1656
## Year2008      -0.60068      0.26335      -2.28      0.0233 *
## Year2009      -0.68195      0.24255      -2.81      0.0053 **
## Year2010      -0.33920      0.24922      -1.36      0.1746
## Year2011      -0.52040      0.28479      -1.83      0.0687 .
## Year2012      -0.58255      0.23303      -2.50      0.0130 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0872, Adjusted R-squared:  0.0134
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 260 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.600  0.876  0.952  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      3.39e-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 1.801 1      1.342
## LastAuthorFemale  1.875 1      1.369
## Year              1.347 16      1.009

```

Residuals from first and last author



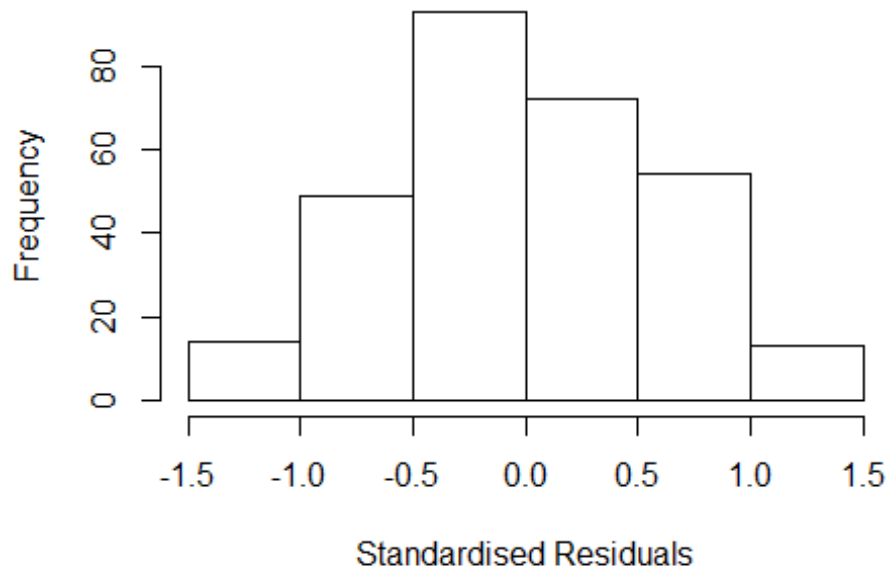
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3198 -0.3814 -0.0116 0.4315 1.4962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4585 0.2252 6.48 4.3e-10 ***
## FirstAuthorFemale1 -0.0405 0.1057 -0.38 0.702
## LastAuthorFemale1 -0.0789 0.1111 -0.71 0.478
## Year1997 -0.4403 0.2975 -1.48 0.140
## Year1998 -0.5069 0.3087 -1.64 0.102
## Year1999 -0.3237 0.2971 -1.09 0.277
## Year2000 -0.4492 0.2886 -1.56 0.121
## Year2001 -0.1387 0.2812 -0.49 0.622
## Year2002 -0.5701 0.2591 -2.20 0.029 *
## Year2003 -0.4891 0.2615 -1.87 0.063 .
## Year2004 -0.4211 0.2602 -1.62 0.107
## Year2005 -0.3253 0.2576 -1.26 0.208
```

```

## Year2006          -0.4239      0.2526   -1.68    0.094 .
## Year2007          -0.3996      0.2651   -1.51    0.133
## Year2008          -0.6890      0.2736   -2.52    0.012 *
## Year2009          -0.6705      0.2664   -2.52    0.012 *
## Year2010          -0.3613      0.2680   -1.35    0.179
## Year2011          -0.5657      0.2895   -1.95    0.052 .
## Year2012          -0.6171      0.2483   -2.48    0.014 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.0612, Adjusted R-squared:  -4.73e-05
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 263 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.570  0.874  0.953  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      3.39e-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 1.142 1      1.069
## Year              1.142 16      1.004

```

Residuals from first author



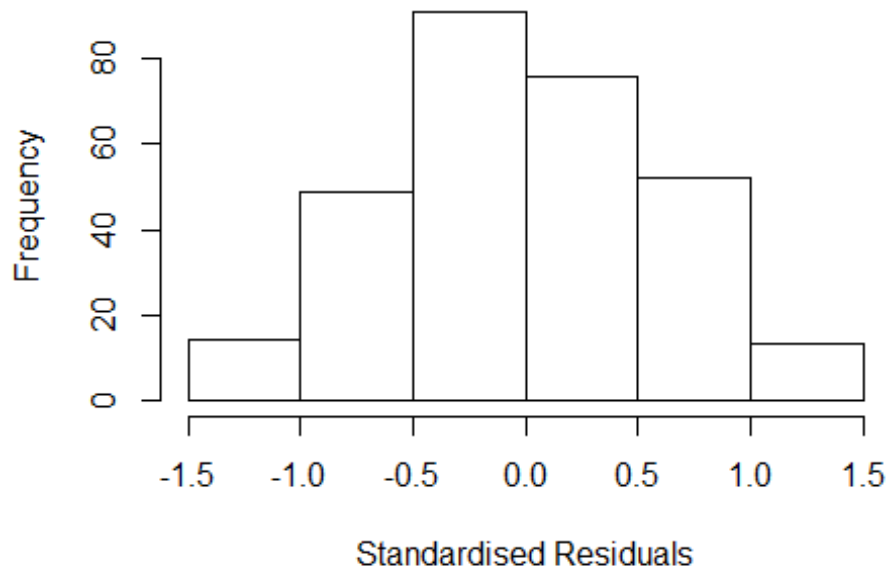
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3216 -0.3863 -0.0203 0.4333 1.4944
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4586 0.2254 6.47 4.4e-10 ***
## FirstAuthorFemale1 -0.0807 0.0855 -0.94 0.346
## Year1997 -0.4482 0.2981 -1.50 0.134
## Year1998 -0.5131 0.3104 -1.65 0.099 .
## Year1999 -0.3279 0.2975 -1.10 0.271
## Year2000 -0.4555 0.2866 -1.59 0.113
## Year2001 -0.1370 0.2824 -0.49 0.628
## Year2002 -0.5656 0.2591 -2.18 0.030 *
## Year2003 -0.5010 0.2620 -1.91 0.057 .
## Year2004 -0.4209 0.2605 -1.62 0.107
## Year2005 -0.3344 0.2574 -1.30 0.195
## Year2006 -0.4326 0.2516 -1.72 0.087 .
```

```

## Year2007          -0.4023      0.2651   -1.52    0.130
## Year2008          -0.6890      0.2738   -2.52    0.012 *
## Year2009          -0.6705      0.2666   -2.52    0.012 *
## Year2010          -0.3726      0.2676   -1.39    0.165
## Year2011          -0.5637      0.2895   -1.95    0.053 .
## Year2012          -0.6261      0.2482   -2.52    0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.0601, Adjusted R-squared:  0.00241
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 264 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.568  0.869  0.953  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      3.39e-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 1.21 1          1.100
## Year              1.21 16          1.006

```

Residuals from last author



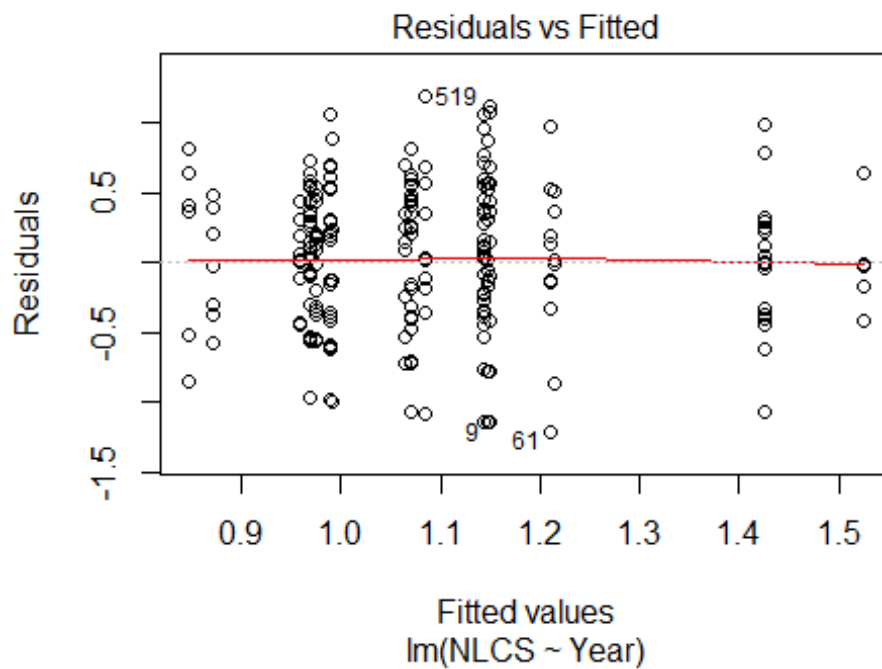
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3131 -0.4015 -0.0114 0.4322 1.5029
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4586 0.2254 6.47 4.4e-10 ***
## LastAuthorFemale1 -0.1025 0.0895 -1.15 0.253
## Year1997 -0.4436 0.2978 -1.49 0.137
## Year1998 -0.5097 0.3096 -1.65 0.101
## Year1999 -0.3256 0.2974 -1.09 0.274
## Year2000 -0.4495 0.2896 -1.55 0.122
## Year2001 -0.1455 0.2791 -0.52 0.603
## Year2002 -0.5756 0.2584 -2.23 0.027 *
## Year2003 -0.4944 0.2616 -1.89 0.060 .
## Year2004 -0.4255 0.2603 -1.63 0.103
## Year2005 -0.3284 0.2577 -1.27 0.204
## Year2006 -0.4256 0.2528 -1.68 0.093 .
```

```

## Year2007          -0.4046      0.2646    -1.53      0.127
## Year2008          -0.6890      0.2739    -2.52      0.012 *
## Year2009          -0.6705      0.2666    -2.51      0.012 *
## Year2010          -0.3636      0.2678    -1.36      0.176
## Year2011          -0.5688      0.2899    -1.96      0.051 .
## Year2012          -0.6194      0.2482    -2.50      0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.0609, Adjusted R-squared:  0.0033
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 262 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.563  0.873   0.954   0.917   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      3.39e-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: 295"
## [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
##   35   31   21   21   36   26   14   19   23   23   37   46   23   46   47
## 2011 2012
##   39   48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13    8    5    7    9    5    4    8   10    8   18   25   11   25   30
## 2011 2012

```

```
## 16 30
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 7 5 6 6 4 3 8 9 5 18 21 10 19 27
## 2011 2012
## 16 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 = 19, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.6, df = 1, p-value = 0.4

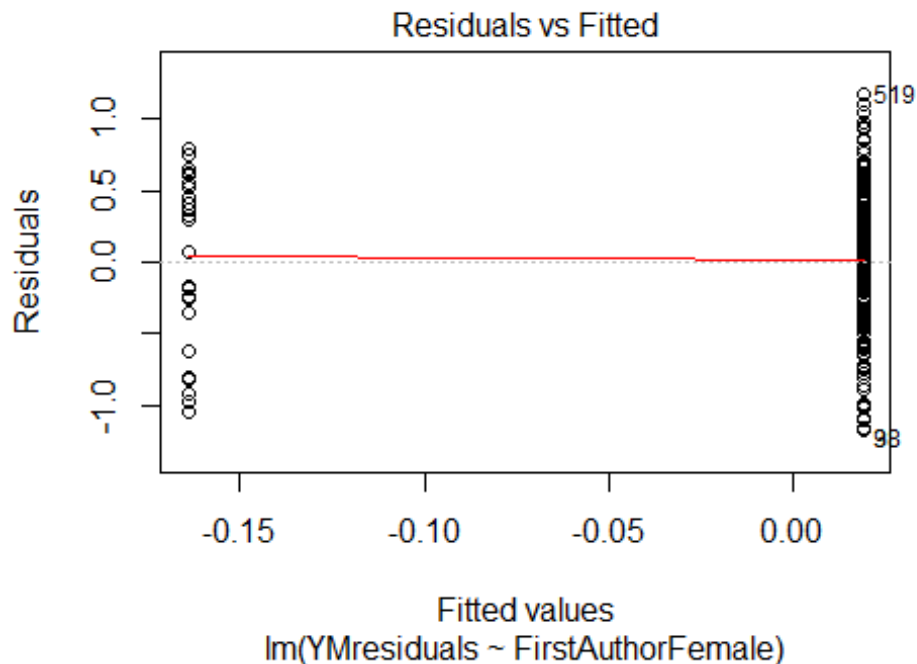
## [1] "Female first author team size 2018 geometric mean: 2"
## [1] "Male first author team size 2018 geometric mean: 1.27917542745868"

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



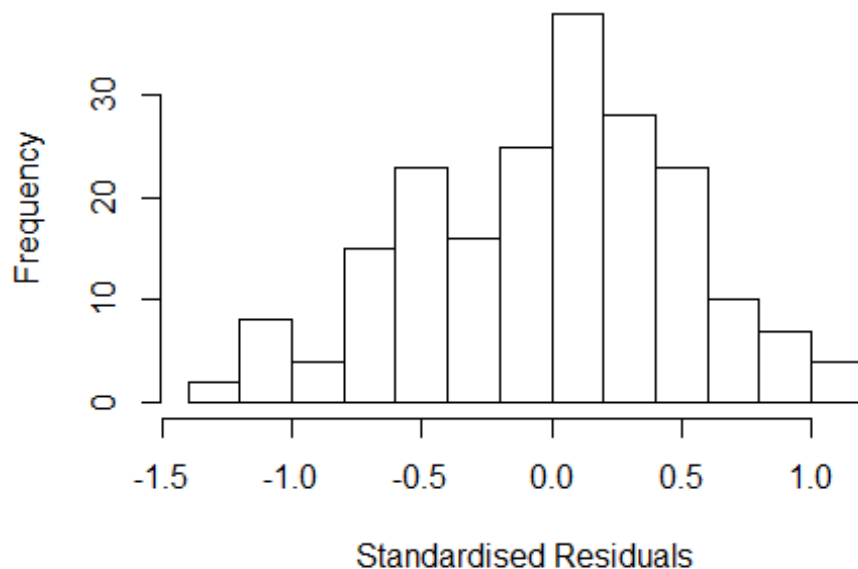
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 50, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1"
## [1] "Male last author team size 2018 geometric mean: 1.36301080612251"

## 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 = 8, 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 1.634 1      1.278
## LastAuthorFemale  1.675 1      1.294
## UniqueAuthors    2.299 3      1.149
## Year              4.066 16     1.045
```

Residuals from first and last author and team size



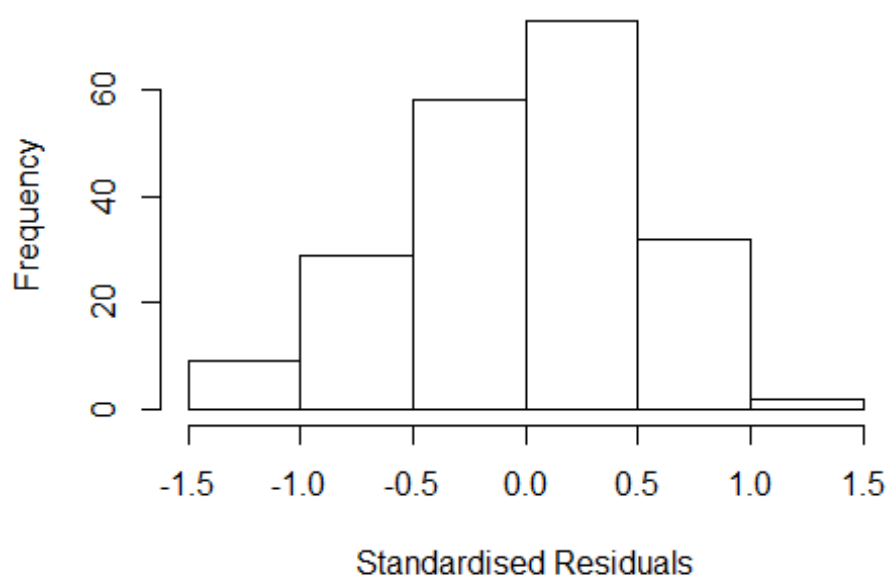
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3429 -0.4057 0.0422 0.3523 1.1039
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3429 0.2863 4.69 5.4e-06 ***
## FirstAuthorFemale1 -0.2449 0.1739 -1.41 0.1608
## LastAuthorFemale1 -0.1375 0.2191 -0.63 0.5309
## UniqueAuthors2 0.2700 0.0945 2.86 0.0048 **
## UniqueAuthors3 -0.0959 0.1486 -0.65 0.5194
## UniqueAuthors4 0.1918 0.2534 0.76 0.4501
## Year1997 -0.0475 0.3416 -0.14 0.8896
## Year1998 -0.2476 0.4060 -0.61 0.5428
## Year1999 -0.2755 0.4170 -0.66 0.5096
## Year2000 -0.4936 0.3060 -1.61 0.1085
```

```

## Year2001          0.0245      0.2978      0.08      0.9344
## Year2002         -0.7890      0.3605     -2.19      0.0299 *
## Year2003         -0.4451      0.3325     -1.34      0.1823
## Year2004         -0.2498      0.3394     -0.74      0.4627
## Year2005         -0.1994      0.3701     -0.54      0.5907
## Year2006          0.0589      0.3024      0.19      0.8457
## Year2007         -0.2805      0.3033     -0.92      0.3562
## Year2008         -0.1706      0.3311     -0.52      0.6070
## Year2009         -0.5486      0.3153     -1.74      0.0836 .
## Year2010         -0.2388      0.2998     -0.80      0.4267
## Year2011         -0.4185      0.2931     -1.43      0.1551
## Year2012         -0.4574      0.3014     -1.52      0.1309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.169, Adjusted R-squared:  0.0727
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 184 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.477  0.875   0.943   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      4.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.466 1      1.211
## LastAuthorFemale  1.630 1      1.277
## Year              1.839 16      1.019

```

Residuals from first and last author



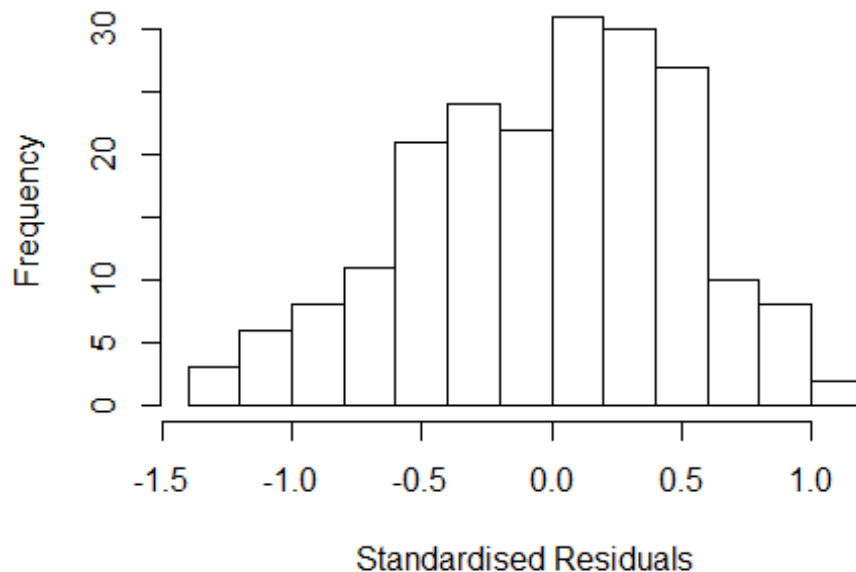
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3877 -0.3932  0.0552  0.3449  1.2378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.38766    0.30569   4.54    1e-05 ***
## FirstAuthorFemale1 -0.19934    0.18719  -1.06    0.29
## LastAuthorFemale1 -0.16621    0.23015  -0.72    0.47
## Year1997          -0.00495    0.35371  -0.01    0.99
## Year1998          -0.13837    0.37370  -0.37    0.71
## Year1999          -0.31292    0.39122  -0.80    0.42
## Year2000          -0.47846    0.32817  -1.46    0.15
## Year2001          -0.02032    0.31651  -0.06    0.95
## Year2002          -0.66469    0.45492  -1.46    0.15
## Year2003          -0.46008    0.34263  -1.34    0.18
## Year2004          -0.27482    0.36974  -0.74    0.46
## Year2005          -0.18067    0.36991  -0.49    0.63
```

```

## Year2006          0.05347      0.32425      0.16      0.87
## Year2007          -0.28635      0.32229     -0.89      0.38
## Year2008          -0.18520      0.36736     -0.50      0.61
## Year2009          -0.50445      0.33357     -1.51      0.13
## Year2010          -0.21131      0.32406     -0.65      0.52
## Year2011          -0.39688      0.31278     -1.27      0.21
## Year2012          -0.39638      0.32278     -1.23      0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.123, Adjusted R-squared:  0.0369
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 188 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.486  0.890  0.947  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      4.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.157 1      1.076
## Year              1.157 16      1.005

```

Residuals from first author



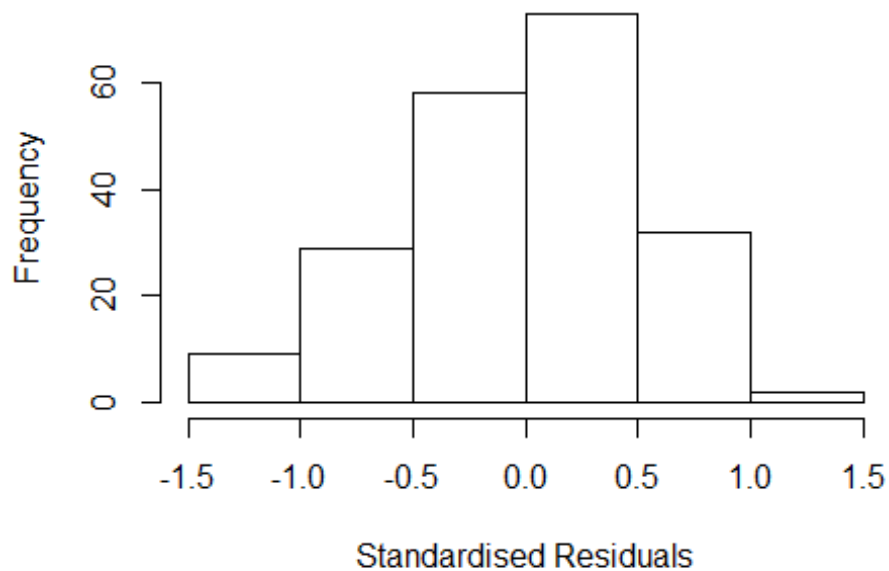
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.386 -0.395 0.037 0.354 1.071
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.38599 0.30926 4.48 1.3e-05 ***
## FirstAuthorFemale1 -0.28428 0.18472 -1.54 0.13
## Year1997 -0.00571 0.37138 -0.02 0.99
## Year1998 -0.13826 0.37726 -0.37 0.71
## Year1999 -0.29895 0.38809 -0.77 0.44
## Year2000 -0.47691 0.33160 -1.44 0.15
## Year2001 -0.01869 0.32000 -0.06 0.95
## Year2002 -0.66419 0.45641 -1.46 0.15
## Year2003 -0.48222 0.34483 -1.40 0.16
## Year2004 -0.27431 0.37278 -0.74 0.46
## Year2005 -0.21587 0.37356 -0.58 0.56
## Year2006 0.05973 0.32920 0.18 0.86
```

```

## Year2007      -0.29170    0.32572   -0.90    0.37
## Year2008      -0.18299    0.36612   -0.50    0.62
## Year2009      -0.53187    0.33811   -1.57    0.12
## Year2010      -0.21002    0.32761   -0.64    0.52
## Year2011      -0.38953    0.31778   -1.23    0.22
## Year2012      -0.40469    0.32687   -1.24    0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.118, Adjusted R-squared:  0.0369
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 186 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.504  0.886  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      4.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.35 1          1.162
## Year              1.35 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
## -1.3492 -0.3823 0.0497 0.3458 1.3637
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3492 0.2995 4.50 1.2e-05 ***
## LastAuthorFemale1 -0.2875 0.2249 -1.28 0.20
## Year1997 -0.0198 0.3630 -0.05 0.96
## Year1998 -0.0994 0.3693 -0.27 0.79
## Year1999 -0.3040 0.4143 -0.73 0.46
## Year2000 -0.4399 0.3226 -1.36 0.17
## Year2001 0.0182 0.3106 0.06 0.95
## Year2002 -0.6258 0.4518 -1.39 0.17
## Year2003 -0.4126 0.3293 -1.25 0.21
## Year2004 -0.2360 0.3650 -0.65 0.52
## Year2005 -0.1143 0.3542 -0.32 0.75
## Year2006 0.0796 0.3195 0.25 0.80
```

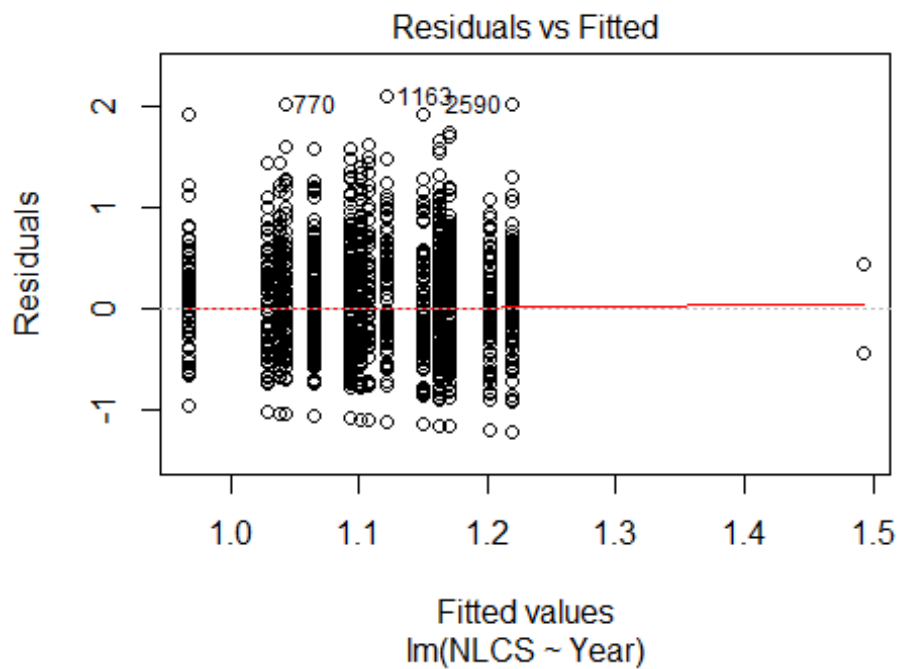


```

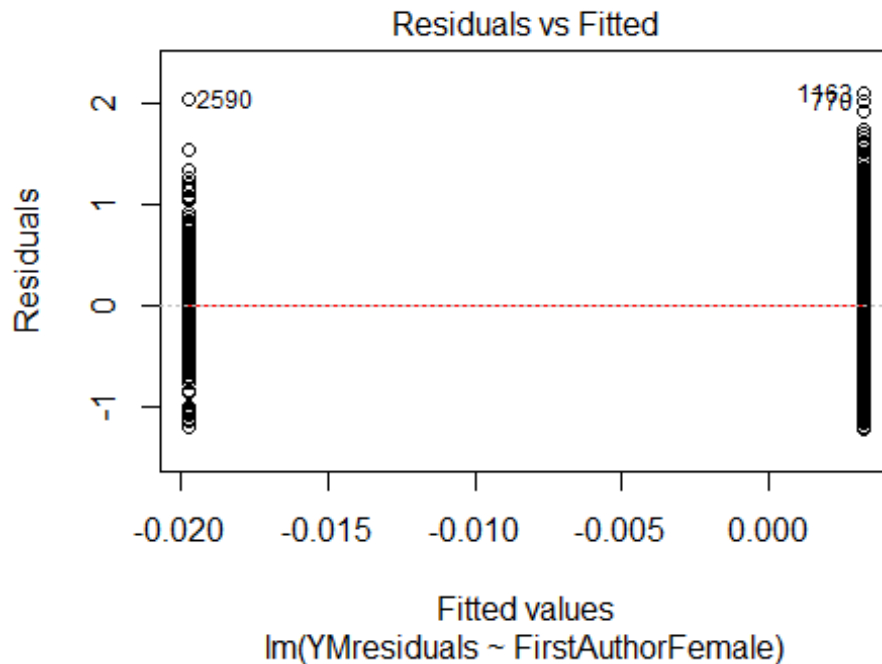
## Year2007          -0.2538      0.3187    -0.80      0.43
## Year2008          -0.1514      0.3691    -0.41      0.68
## Year2009          -0.4633      0.3389    -1.37      0.17
## Year2010          -0.1859      0.3202    -0.58      0.56
## Year2011          -0.3714      0.3119    -1.19      0.24
## Year2012          -0.3600      0.3162    -1.14      0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.116, Adjusted R-squared:  0.0344
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 185 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.496  0.881  0.948  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      4.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 203"
## [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
## 235 246 215 12 21 252 194 172 159 198 234 240 268 295 274
## 2011 2012
## 292 304
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 87 74 2 7 96 81 80 67 110 121 112 144 153 152
## 2011 2012

```

```
## 151 165
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 66 72 65 2 4 82 68 69 60 93 104 96 124 125 120
## 2011 2012
## 130 144
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 37, df = 16, p-value = 0.002
```

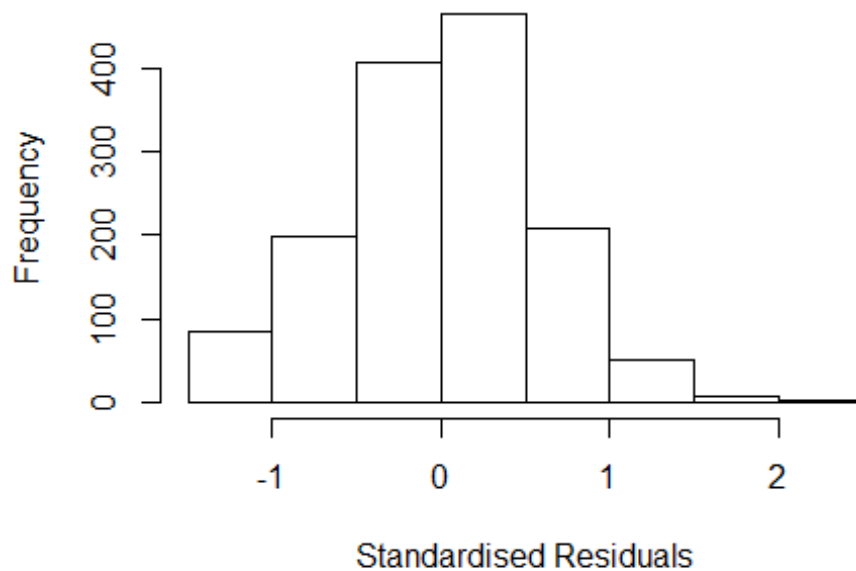


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.17, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 2.86672610948252"
## [1] "Male first author team size 2018 geometric mean: 1.86604217536026"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.008
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.56594069765072"
## [1] "Male last author team size 2018 geometric mean: 1.89472492825527"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, 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 1.305 1          1.142
## LastAuthorFemale  1.377 1          1.174
## UniqueAuthors     1.437 4          1.046
## Year              1.559 16         1.014
```

Residuals from first and last author and team size



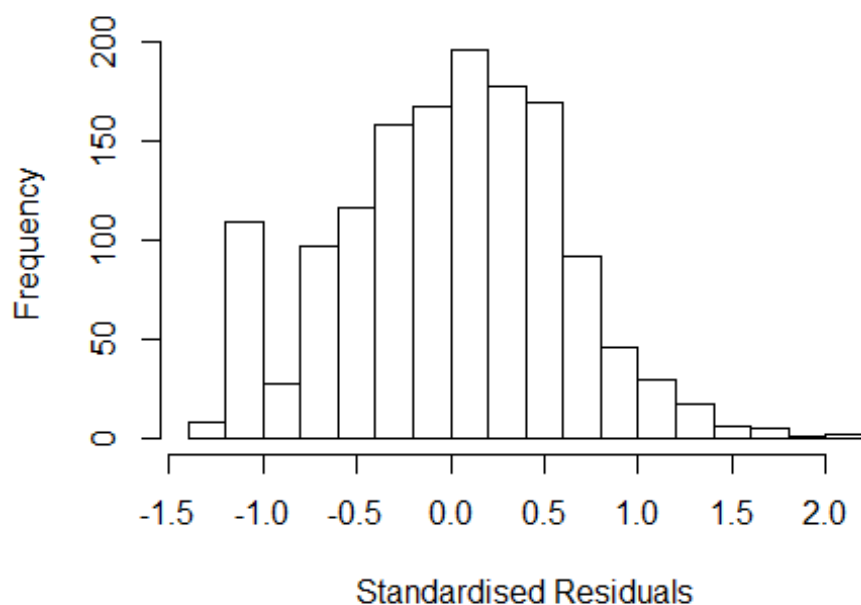
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3476 -0.3971 0.0258 0.3894 2.1486
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05069 0.10364 10.14 < 2e-16 ***
## FirstAuthorFemale1 -0.01746 0.05106 -0.34 0.73
## LastAuthorFemale1 -0.07148 0.05661 -1.26 0.21
## UniqueAuthors2 0.20628 0.03828 5.39 8.3e-08 ***
## UniqueAuthors3 0.19539 0.04765 4.10 4.4e-05 ***
## UniqueAuthors4 0.29948 0.07379 4.06 5.2e-05 ***
## UniqueAuthors5 0.21767 0.10000 2.18 0.03 *
## Year1997 -0.03266 0.12641 -0.26 0.80
## Year1998 -0.12633 0.14224 -0.89 0.37
## Year1999 0.24098 0.36384 0.66 0.51
```

```

## Year2000      0.13716    0.10938    1.25    0.21
## Year2001     -0.00259    0.13585   -0.02    0.98
## Year2002     -0.18573    0.13080   -1.42    0.16
## Year2003     -0.15513    0.12309   -1.26    0.21
## Year2004      0.08123    0.12515    0.65    0.52
## Year2005     -0.07828    0.11738   -0.67    0.50
## Year2006      0.05047    0.11571    0.44    0.66
## Year2007     -0.00798    0.11710   -0.07    0.95
## Year2008     -0.14234    0.11106   -1.28    0.20
## Year2009     -0.09841    0.11512   -0.85    0.39
## Year2010     -0.02371    0.11378   -0.21    0.83
## Year2011     -0.13842    0.11056   -1.25    0.21
## Year2012     -0.11734    0.11278   -1.04    0.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.0448, Adjusted R-squared:  0.0298
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1303 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.144  0.868  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      7.02e-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 1.262 1      1.123
## LastAuthorFemale 1.301 1      1.140
## Year      1.140 16      1.004

```

Residuals from first and last author

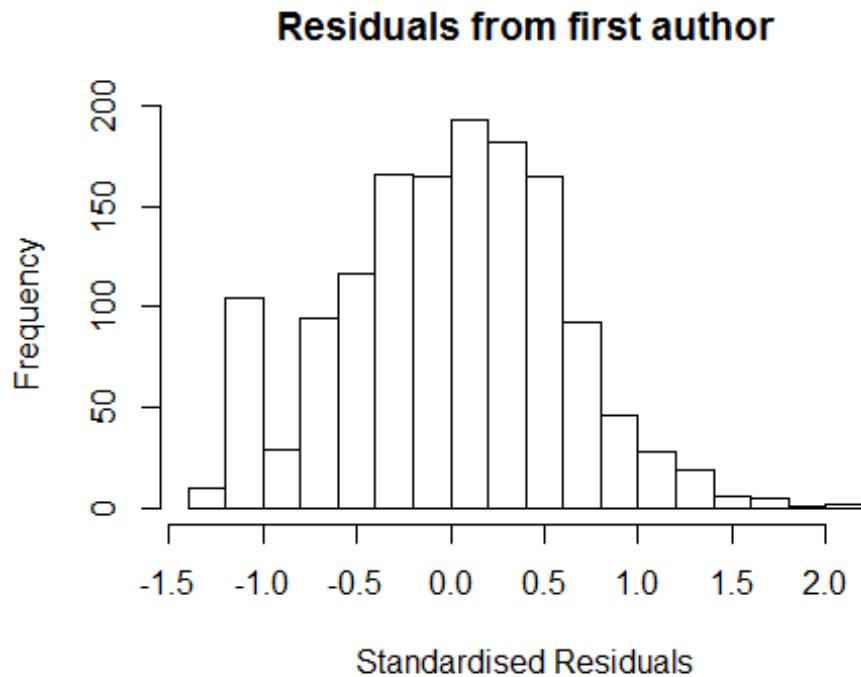


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.225 -0.403 0.032 0.413 2.074
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15468 0.10173 11.35 <2e-16 ***
## FirstAuthorFemale1 0.00318 0.05138 0.06 0.95
## LastAuthorFemale1 -0.06326 0.05661 -1.12 0.26
## Year1997 -0.06979 0.12794 -0.55 0.59
## Year1998 -0.11976 0.14136 -0.85 0.40
## Year1999 0.33782 0.35742 0.95 0.34
## Year2000 0.07769 0.11704 0.66 0.51
## Year2001 -0.01587 0.13500 -0.12 0.91
## Year2002 -0.16056 0.13036 -1.23 0.22
## Year2003 -0.15409 0.12195 -1.26 0.21
## Year2004 0.06775 0.12645 0.54 0.59
## Year2005 -0.06447 0.11851 -0.54 0.59
```

```

## Year2006          0.07002      0.11590      0.60      0.55
## Year2007          0.00738      0.11935      0.06      0.95
## Year2008         -0.11136      0.11205     -0.99      0.32
## Year2009         -0.07732      0.11559     -0.67      0.50
## Year2010         -0.00869      0.11475     -0.08      0.94
## Year2011         -0.11577      0.11115     -1.04      0.30
## Year2012         -0.07203      0.11390     -0.63      0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.0142, Adjusted R-squared:  0.0016
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1306 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.194  0.876  0.950  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      7.02e-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 1.065 1      1.032
## Year              1.065 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
## -1.2236 -0.3960  0.0328  0.4146  2.0414
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14945    0.10245   11.22  <2e-16 ***
## FirstAuthorFemale1 -0.01978    0.04749   -0.42    0.68
## Year1997        -0.07028    0.12847   -0.55    0.58
## Year1998        -0.11788    0.14167   -0.83    0.41
## Year1999         0.34305    0.35701    0.96    0.34
## Year2000         0.07298    0.12137    0.60    0.55
## Year2001        -0.01171    0.13537   -0.09    0.93
## Year2002        -0.16250    0.13036   -1.25    0.21
## Year2003        -0.15638    0.12261   -1.28    0.20
## Year2004         0.06606    0.12725    0.52    0.60
## Year2005        -0.05994    0.11919   -0.50    0.62
## Year2006         0.07412    0.11658    0.64    0.53
```

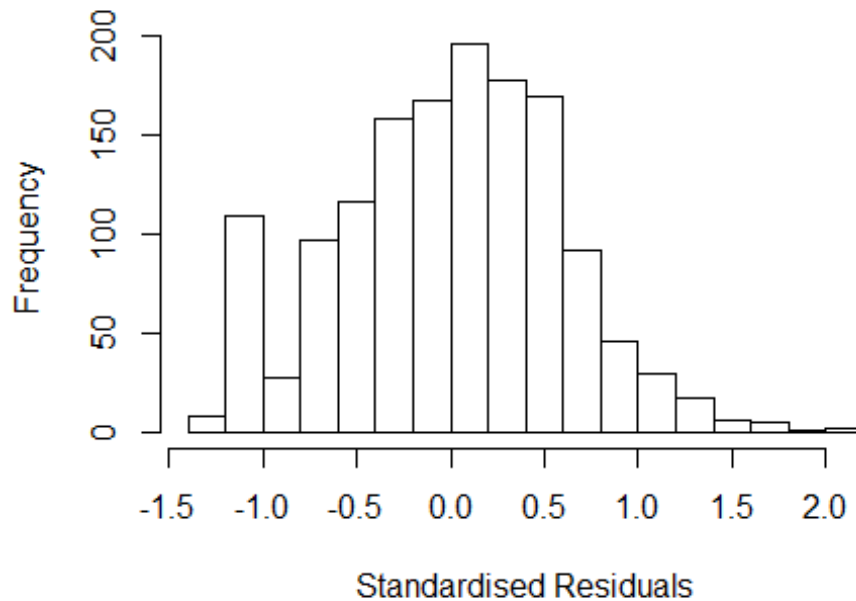


```

## Year2007          0.01142    0.11967    0.10    0.92
## Year2008          -0.10861    0.11276   -0.96    0.34
## Year2009          -0.08206    0.11602   -0.71    0.48
## Year2010          -0.00843    0.11541   -0.07    0.94
## Year2011          -0.11636    0.11186   -1.04    0.30
## Year2012          -0.07209    0.11463   -0.63    0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.0133, Adjusted R-squared:  0.00142
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1302 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.217  0.876  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      7.02e-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 1.096 1          1.047
## Year            1.096 16          1.003

```

Residuals from last author



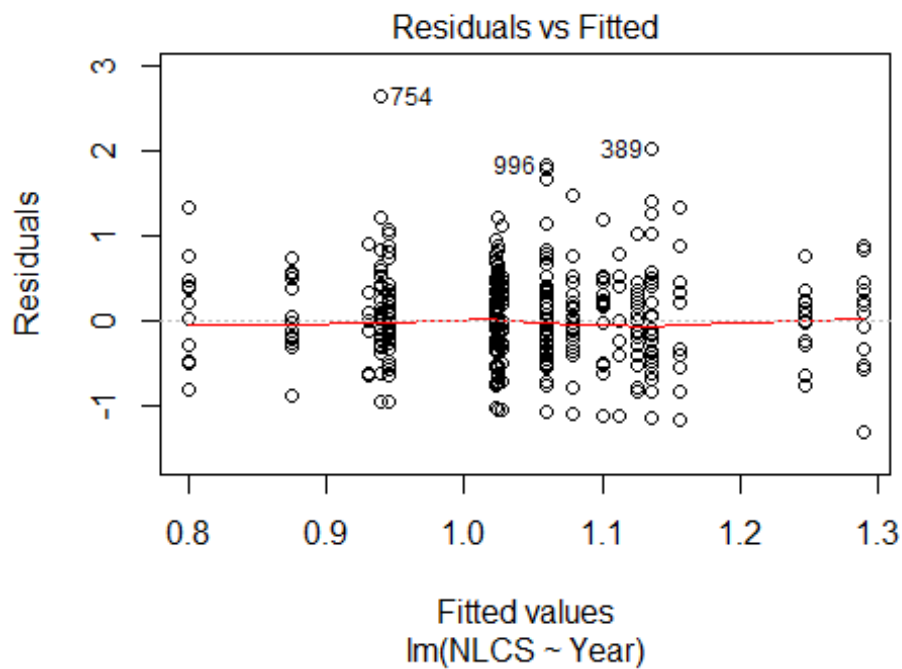
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2249 -0.4028 0.0317 0.4127 2.0761
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15498 0.10162 11.37 <2e-16 ***
## LastAuthorFemale1 -0.06197 0.05201 -1.19 0.23
## Year1997 -0.06977 0.12802 -0.55 0.59
## Year1998 -0.11977 0.14137 -0.85 0.40
## Year1999 0.33752 0.35746 0.94 0.35
## Year2000 0.07786 0.11694 0.67 0.51
## Year2001 -0.01588 0.13508 -0.12 0.91
## Year2002 -0.16061 0.13042 -1.23 0.22
## Year2003 -0.15397 0.12196 -1.26 0.21
## Year2004 0.06787 0.12654 0.54 0.59
## Year2005 -0.06463 0.11852 -0.55 0.59
## Year2006 0.06989 0.11588 0.60 0.55
```

```

## Year2007      0.00766    0.11937    0.06    0.95
## Year2008     -0.11125    0.11209   -0.99    0.32
## Year2009     -0.07730    0.11563   -0.67    0.50
## Year2010     -0.00882    0.11475   -0.08    0.94
## Year2011     -0.11583    0.11118   -1.04    0.30
## Year2012     -0.07212    0.11392   -0.63    0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0142, Adjusted R-squared:  0.00232
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1305 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.875  0.950  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      7.02e-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: 1424"
## [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
##   29   33   32   37   51   35   41   39   55   47   51   46   73   59   68
## 2011 2012
##   85   87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14    7   10   13    7   13   19   14   28   25   26   15   41   32   34
## 2011 2012

```

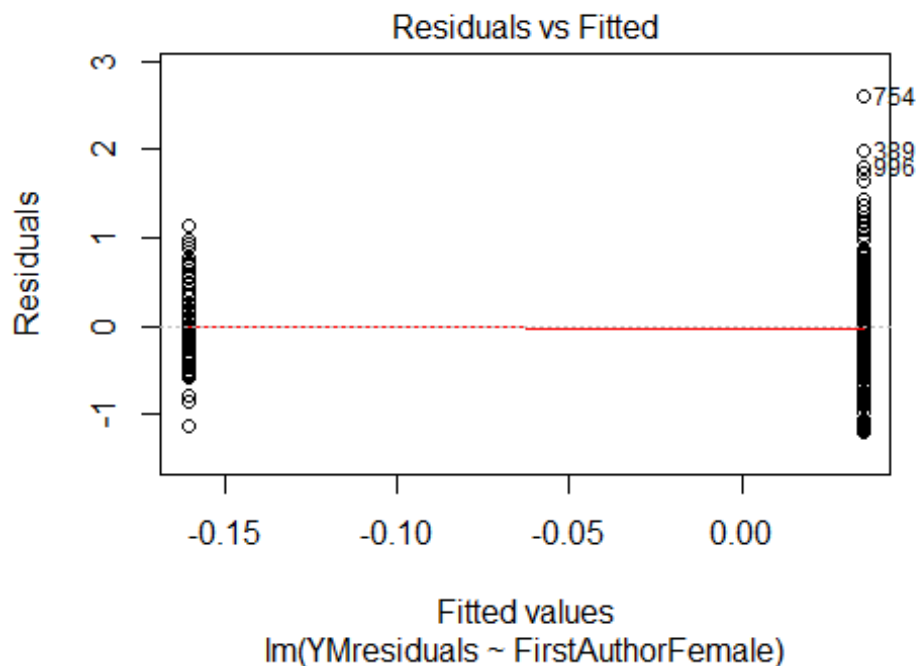
```
## 48 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 6 9 10 6 11 17 12 26 24 20 11 36 29 29
## 2011 2012
## 39 45
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, df = 16, p-value = 0.2
```



```
##
## 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: 6.05480020807018"
## [1] "Male first author team size 2018 geometric mean: 1.6971012317068"
## 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 = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3"
## [1] "Male last author team size 2018 geometric mean: 1.99427343705643"

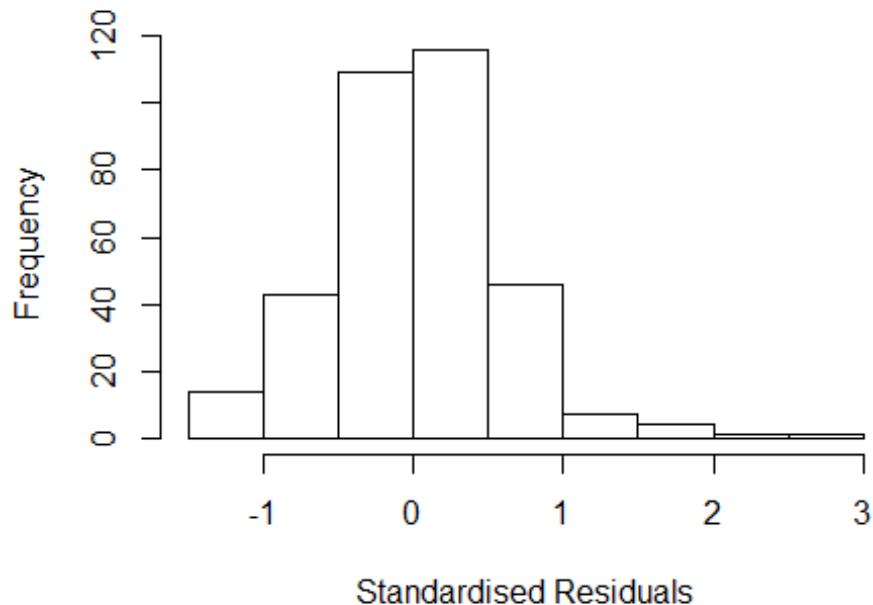
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20, p-value = 0.5
## 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	1.550	1	1.245
LastAuthorFemale	1.526	1	1.235
UniqueAuthors	2.948	4	1.145
Year	4.814	16	1.050

Residuals from first and last author and team size



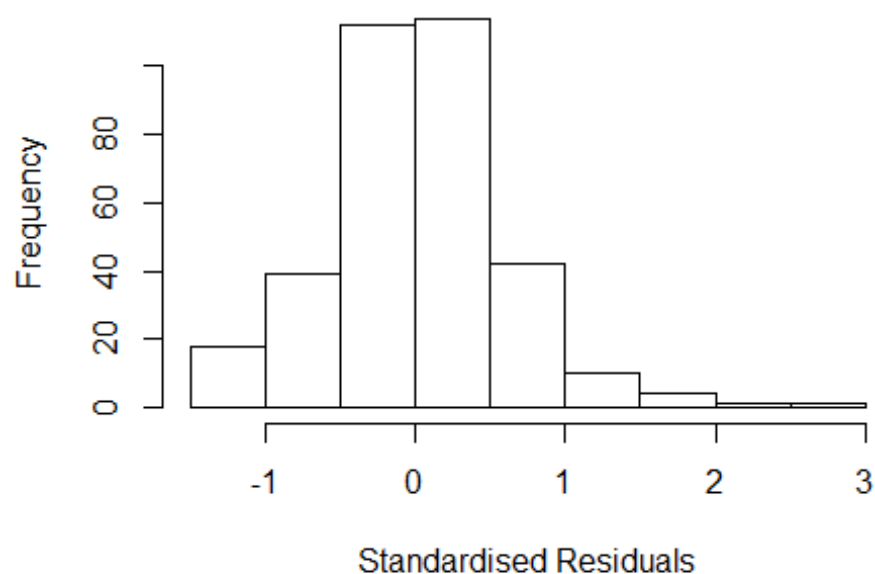
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 75249087100 3.574 2009      1303      6      2.652
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26276 -0.32810  0.00682  0.35533  2.65192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.81751    0.22937   3.56 0.00042 ***
## FirstAuthorFemale1 -0.23259    0.07660  -3.04 0.00259 **
## LastAuthorFemale1  0.10565    0.08461   1.25 0.21272
## UniqueAuthors2    0.11620    0.09081   1.28 0.20164
## UniqueAuthors3    0.09641    0.08682   1.11 0.26763
## UniqueAuthors4    0.19068    0.11787   1.62 0.10672
## UniqueAuthors5    0.34037    0.09794   3.48 0.00058 ***
## Year1997        -0.00800    0.30971  -0.03 0.97941
## Year1998         0.30895    0.37451   0.82 0.41002
## Year1999         0.42548    0.22671   1.88 0.06147 .
```

```

## Year2000      0.15160      0.31071      0.49  0.62595
## Year2001      0.48717      0.27464      1.77  0.07705 .
## Year2002      0.11839      0.22749      0.52  0.60313
## Year2003      0.01307      0.26581      0.05  0.96081
## Year2004      0.10136      0.25270      0.40  0.68862
## Year2005      0.16410      0.23180      0.71  0.47949
## Year2006      0.20018      0.26112      0.77  0.44388
## Year2007      0.19345      0.26703      0.72  0.46931
## Year2008      0.07019      0.22867      0.31  0.75908
## Year2009      0.00817      0.23644      0.03  0.97247
## Year2010      0.11133      0.25442      0.44  0.66198
## Year2011      0.08732      0.23338      0.37  0.70855
## Year2012      0.13586      0.22546      0.60  0.54722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.089, Adjusted R-squared:  0.026
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 205 is an outlier with |weight| = 0 ( < 0.00029);
## 33 weights are ~= 1. The remaining 307 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0208 0.8620 0.9510 0.8930 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      2.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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 1.415 1      1.189
## Year      1.806 16      1.019

```

Residuals from first and last author



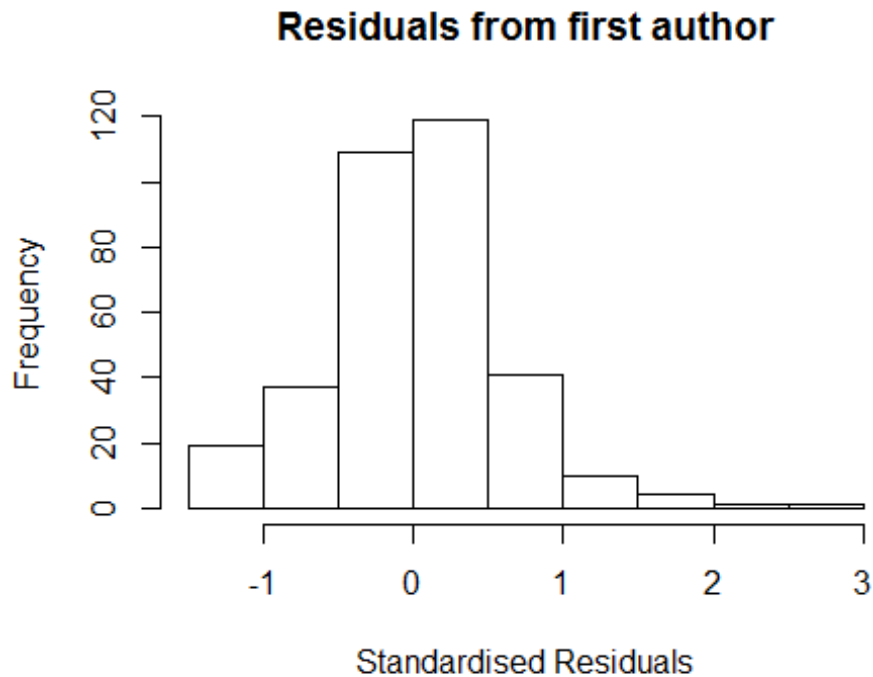
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 75249087100 3.574 2009      1303      6      2.64
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2095 -0.3326  0.0226  0.3343  2.6399
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9151     0.2013   4.55 7.8e-06 ***
## FirstAuthorFemale1 -0.2202     0.0777  -2.83  0.0049 **
## LastAuthorFemale1  0.1103     0.0862   1.28  0.2019
## Year1997          -0.0805     0.2899  -0.28  0.7815
## Year1998           0.2945     0.3691   0.80  0.4256
## Year1999           0.4155     0.2229   1.86  0.0632 .
## Year2000           0.1501     0.3259   0.46  0.6454
## Year2001           0.4655     0.2744   1.70  0.0907 .
## Year2002           0.1228     0.2163   0.57  0.5708
## Year2003          -0.0267     0.2507  -0.11  0.9154
## Year2004           0.0753     0.2422   0.31  0.7561
## Year2005           0.1436     0.2159   0.67  0.5065
```



```

## Year2006          0.1992      0.2367      0.84      0.4006
## Year2007          0.1472      0.2551      0.58      0.5644
## Year2008          0.0763      0.2176      0.35      0.7260
## Year2009          0.0190      0.2219      0.09      0.9317
## Year2010          0.1754      0.2383      0.74      0.4621
## Year2011          0.1169      0.2176      0.54      0.5916
## Year2012          0.1656      0.2149      0.77      0.4414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0596, Adjusted R-squared:  0.00705
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 205 is an outlier with |weight| = 0 ( < 0.00029);
## 28 weights are ~= 1. The remaining 312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0441 0.8570 0.9560 0.8930 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          2.93e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.312 1          1.146
## Year              1.312 16          1.009

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 75249087100 3.574 2009      1303      6      2.64
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2273 -0.3462  0.0205  0.3264  2.6192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95610    0.19929   4.80 2.5e-06 ***
## FirstAuthorFemale1 -0.20443    0.07527  -2.72  0.007 **
## Year1997      -0.11561    0.29512  -0.39  0.696
## Year1998       0.27123    0.37872   0.72  0.474
## Year1999       0.39699    0.22838   1.74  0.083 .
## Year2000       0.10451    0.32133   0.33  0.745
## Year2001       0.42010    0.27095   1.55  0.122
## Year2002       0.11214    0.22253   0.50  0.615
## Year2003      -0.05300    0.24839  -0.21  0.831
## Year2004       0.07505    0.24741   0.30  0.762
## Year2005       0.11746    0.21837   0.54  0.591
## Year2006       0.16029    0.23558   0.68  0.497
```

```

## Year2007          0.11142      0.24935      0.45      0.655
## Year2008          0.04084      0.21590      0.19      0.850
## Year2009         -0.00127      0.22399     -0.01      0.995
## Year2010          0.13168      0.23470      0.56      0.575
## Year2011          0.09208      0.22057      0.42      0.677
## Year2012          0.14280      0.21720      0.66      0.511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0542, Adjusted R-squared:  0.00447
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 205 is an outlier with |weight| = 0 ( < 0.00029);
## 29 weights are ~= 1. The remaining 311 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.065  0.853  0.956  0.895  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      2.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.332 1          1.154
## Year              1.332 16          1.009
##
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 75249087100 3.574 2009      1303      6      2.64
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max

```

```

## -1.3288 -0.3206 0.0122 0.3522 2.6616
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8954    0.1994    4.49 9.9e-06 ***
## LastAuthorFemale1 0.0731    0.0890    0.82  0.412
## Year1997       -0.0572    0.2918   -0.20  0.845
## Year1998        0.2846    0.3853    0.74  0.461
## Year1999        0.4189    0.2204    1.90  0.058 .
## Year2000        0.1185    0.3128    0.38  0.705
## Year2001        0.4334    0.2714    1.60  0.111
## Year2002        0.1423    0.2127    0.67  0.504
## Year2003       -0.0923    0.2460   -0.38  0.708
## Year2004        0.0390    0.2449    0.16  0.873
## Year2005        0.1484    0.2166    0.69  0.494
## Year2006        0.1784    0.2342    0.76  0.447
## Year2007        0.1254    0.2519    0.50  0.619
## Year2008        0.0489    0.2138    0.23  0.819
## Year2009        0.0170    0.2211    0.08  0.939
## Year2010        0.1583    0.2293    0.69  0.490
## Year2011        0.1144    0.2177    0.53  0.600
## Year2012        0.1422    0.2160    0.66  0.511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0384, Adjusted R-squared:  -0.0122
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 205 is an outlier with |weight| = 0 ( < 0.00029);
## 34 weights are ~= 1. The remaining 306 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0444 0.8560 0.9540 0.8940 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           2.93e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)

```

```

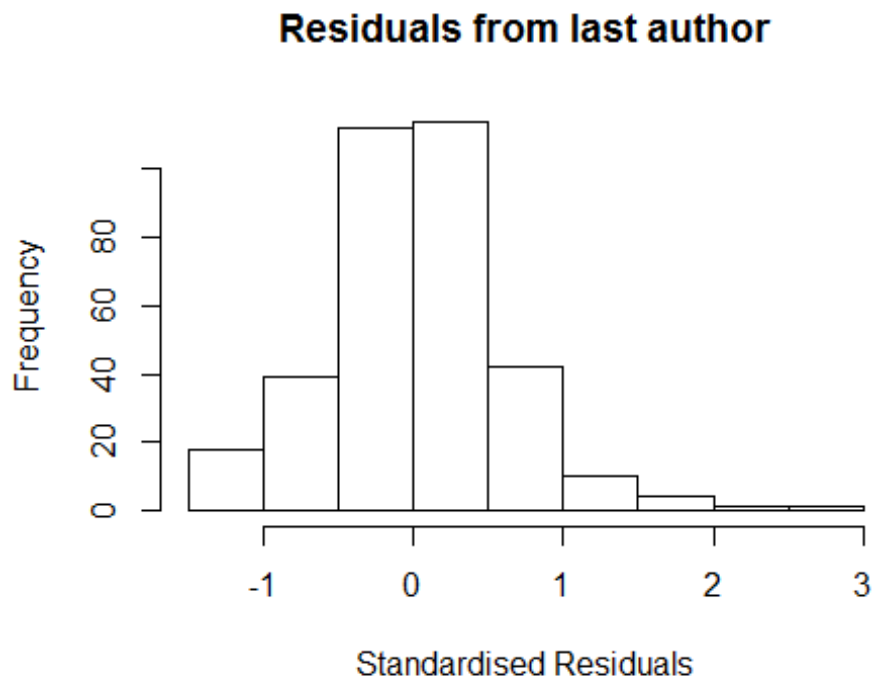
## [1] "Sample size for the above analysis: 341"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2606"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 18 23 17 18 14 13 15 11 15 22 27 29 25 30
## 2011 2012
## 26 35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 0 1 7 4 6 4 6 12 6 10 10 14 8 7 8
## 2011 2012
## 11 22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 0 1 7 4 3 2 5 8 6 8 10 11 6 4 6
## 2011 2012
## 9 16
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.27950705695478"
## [1] "Male first author team size 2018 geometric mean: 2.14697912587862"

## 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 = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.21336383940064"
## [1] "Male last author team size 2018 geometric mean: 2.15515198802076"

## 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 = 0.9
## 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  1.738  1          1.318
## LastAuthorFemale  19.053  1          4.365
## UniqueAuthors    329.801  4          2.064
## Year              603.947 15          1.238

## [1] "List of 0 outliers with 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:
```

```

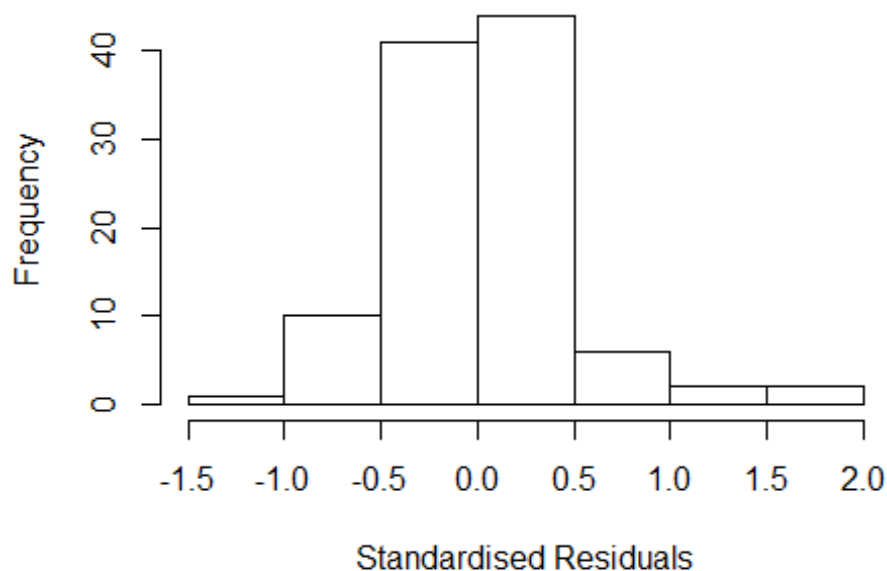
##      Min      1Q  Median      3Q      Max
## -1.0037 -0.2281  0.0248  0.2727  1.8854
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.3859      0.1378    2.80  0.00632 **
## FirstAuthorFemale1 0.2732      0.1262    2.17  0.03318 *
## LastAuthorFemale1 -0.1573      0.2119   -0.74  0.45992
## UniqueAuthors2     0.1521      0.1378    1.10  0.27275
## UniqueAuthors3    -0.0705      0.1325   -0.53  0.59585
## UniqueAuthors4     0.2296      0.1566    1.47  0.14621
## UniqueAuthors5     0.1472      0.1653    0.89  0.37572
## Year1998           0.6859      0.5387    1.27  0.20639
## Year1999           0.8138      0.2264    3.59  0.00055 ***
## Year2000           0.2891      0.1079    2.68  0.00889 **
## Year2001           0.8645      0.0911    9.49  6.3e-15 ***
## Year2002           1.1927      0.1751    6.81  1.4e-09 ***
## Year2003           0.7521      0.1751    4.30  4.7e-05 ***
## Year2004           0.8636      0.1464    5.90  7.5e-08 ***
## Year2005           0.6982      0.1002    6.97  6.8e-10 ***
## Year2006           0.8066      0.1620    4.98  3.4e-06 ***
## Year2007           0.4551      0.1307    3.48  0.00079 ***
## Year2008           0.9485      0.1946    4.88  5.1e-06 ***
## Year2009           0.3536      0.4436    0.80  0.42761
## Year2010          -0.3522      0.1499   -2.35  0.02109 *
## Year2011           0.6178      0.2143    2.88  0.00499 **
## Year2012           0.4691      0.1293    3.63  0.00049 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.435, Adjusted R-squared:  0.294
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## observation 81 is an outlier with |weight| = 0 ( < 0.00094);
## 12 weights are ~= 1. The remaining 93 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0198 0.8720 0.9480 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      9.43e-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"

## 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 1.717 1          1.310
## LastAuthorFemale  3.228 1          1.797
## Year               4.393 15         1.051

## [1] "List of 0 outliers with 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
```



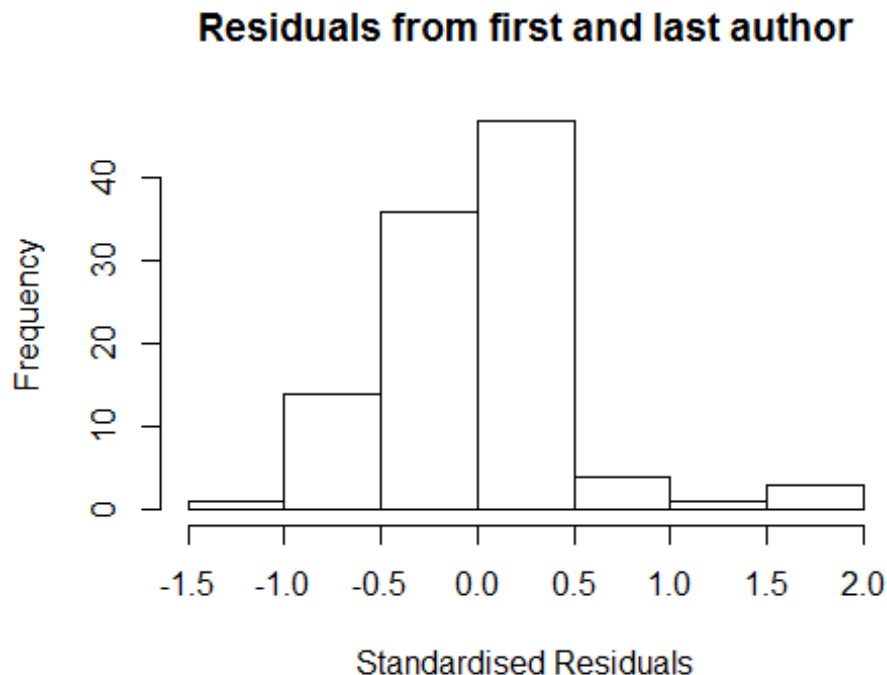
```

## -1.0875 -0.2375 0.0217 0.2948 1.8246
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5380    0.0000      Inf < 2e-16 ***
## FirstAuthorFemale1 0.2516    0.1291     1.95 0.0545 .
## LastAuthorFemale1 -0.1179    0.2750    -0.43 0.6690
## Year1998        0.5064    0.4328     1.17 0.2452
## Year1999        0.6398    0.1938     3.30 0.0014 **
## Year2000        0.1637    0.0164    9.97 4.1e-16 ***
## Year2001        0.8645    0.0912     9.48 4.2e-15 ***
## Year2002        1.0882    0.1388     7.84 9.9e-12 ***
## Year2003        0.6448    0.1567     4.12 8.7e-05 ***
## Year2004        0.7620    0.1467     5.19 1.3e-06 ***
## Year2005        0.6740    0.0911     7.40 7.8e-11 ***
## Year2006        0.7800    0.1754     4.45 2.5e-05 ***
## Year2007        0.3292    0.1016     3.24 0.0017 **
## Year2008        0.7730    0.1636     4.73 8.7e-06 ***
## Year2009        0.1748    0.4294     0.41 0.6849
## Year2010       -0.4612    0.0736    -6.27 1.3e-08 ***
## Year2011        0.5495    0.2461     2.23 0.0281 *
## Year2012        0.3955    0.1357     2.91 0.0045 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared: 0.412, Adjusted R-squared: 0.298
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 81 is an outlier with |weight| <= 0.00031 ( < 0.00094);
## 14 weights are ~= 1. The remaining 91 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0272 0.8680 0.9350 0.8830 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           9.43e-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"

## 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 1.58 1          1.257
## Year              1.58 15         1.015

## [1] "List of 0 outliers 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
## -1.0319 -0.2396  0.0218  0.2958  1.7425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5380    0.0000    Inf < 2e-16 ***
## FirstAuthorFemale1 0.2290    0.1198    1.91  0.05913 .
```

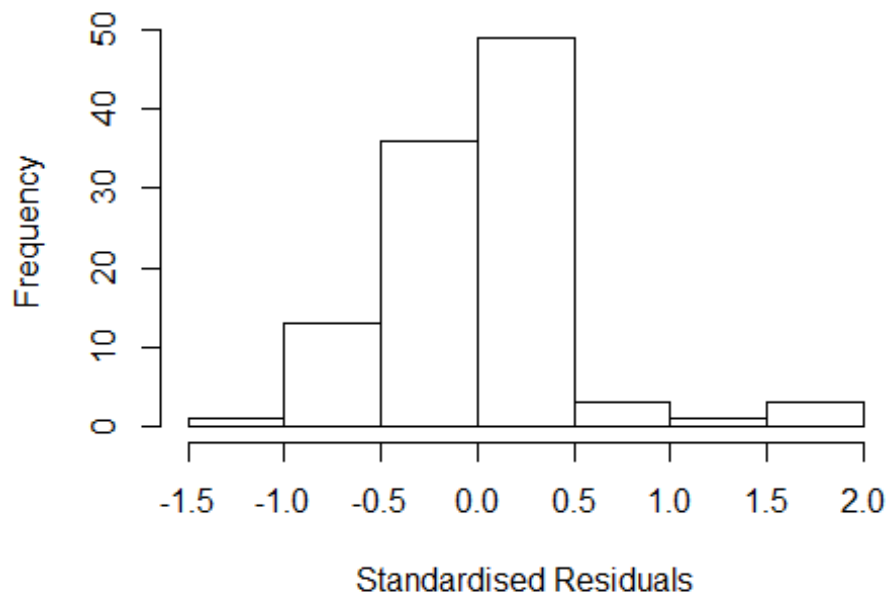
```

## Year1998          0.5056      0.4238      1.19  0.23602
## Year1999          0.6398      0.1938      3.30  0.00138 **
## Year2000          0.1637      0.0164      9.97  3.6e-16 ***
## Year2001          0.8645      0.0912      9.48  3.8e-15 ***
## Year2002          1.0882      0.1388      7.84  9.3e-12 ***
## Year2003          0.6343      0.1705      3.72  0.00035 ***
## Year2004          0.7620      0.1467      5.19  1.3e-06 ***
## Year2005          0.6800      0.0891      7.64  2.4e-11 ***
## Year2006          0.7801      0.1749      4.46  2.4e-05 ***
## Year2007          0.3265      0.1000      3.27  0.00155 **
## Year2008          0.7774      0.1617      4.81  6.1e-06 ***
## Year2009          0.1746      0.4291      0.41  0.68515
## Year2010         -0.4745      0.0634     -7.48  4.9e-11 ***
## Year2011          0.4939      0.1776      2.78  0.00662 **
## Year2012          0.3934      0.1385      2.84  0.00559 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.408, Adjusted R-squared:  0.302
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 91 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0106 0.8670 0.9310 0.8710 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      9.43e-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"

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

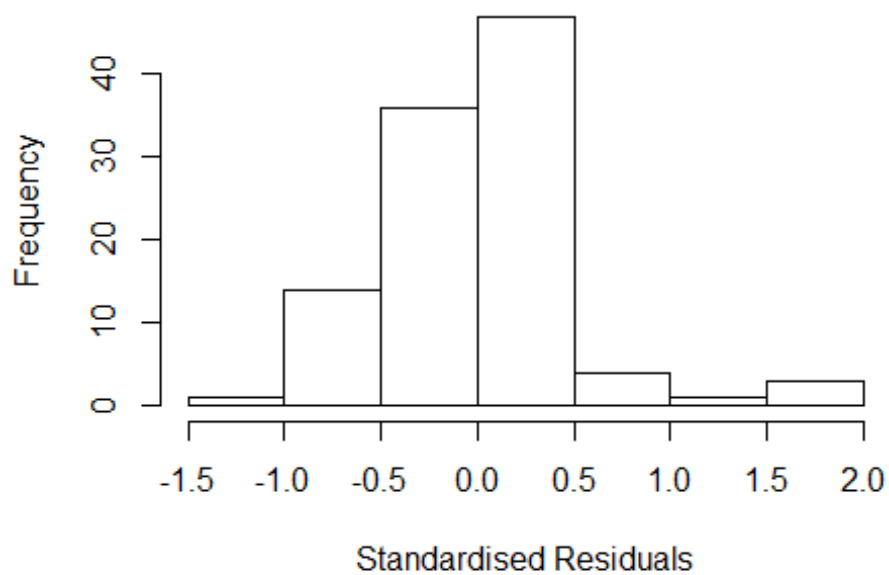
```

Residuals from first author



##	GVIF	Df	$GVIF^{(1/(2*Df))}$
## LastAuthorFemale	3.641	1	1.908
## Year	3.641	15	1.044

Residuals from last author

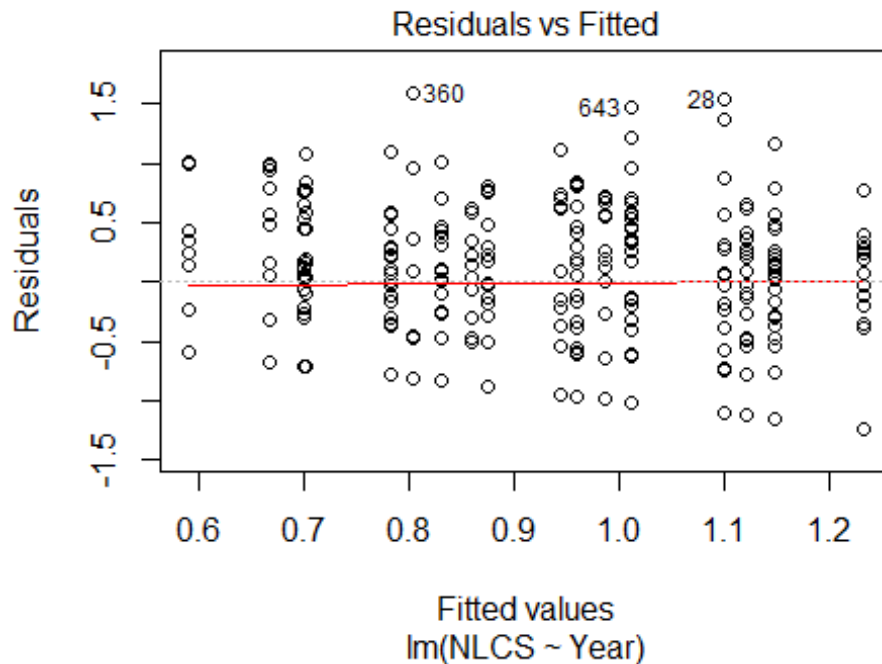


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0819 -0.1662 0.0163 0.2811 1.9583
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5380 0.0000 Inf < 2e-16 ***
## LastAuthorFemale1 -0.0249 0.3033 -0.08 0.93484
## Year1998 0.5064 0.4178 1.21 0.22864
## Year1999 0.6398 0.1937 3.30 0.00138 **
## Year2000 0.1637 0.0164 9.97 3.6e-16 ***
## Year2001 0.8645 0.0912 9.48 3.8e-15 ***
## Year2002 1.0882 0.1388 7.84 9.2e-12 ***
## Year2003 0.6362 0.1648 3.86 0.00021 ***
## Year2004 0.7620 0.1466 5.20 1.3e-06 ***
## Year2005 0.7415 0.0777 9.54 2.8e-15 ***
## Year2006 0.7800 0.1744 4.47 2.3e-05 ***
## Year2007 0.3851 0.1007 3.82 0.00024 ***
## Year2008 0.8237 0.1413 5.83 8.9e-08 ***
## Year2009 0.1748 0.4299 0.41 0.68522
## Year2010 -0.4364 0.0965 -4.52 1.9e-05 ***
## Year2011 0.5439 0.2693 2.02 0.04644 *
## Year2012 0.4402 0.1370 3.21 0.00183 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared: 0.395, Adjusted R-squared: 0.286
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 81 is an outlier with |weight| = 0 ( < 0.00094);
## 15 weights are ~ 1. The remaining 90 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0272 0.8420 0.9410 0.8780 0.9890 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 9.43e-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: 106"
## [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
##    28   34   34   33   37   30   31   23   30   23   35   28   33   44   45
## 2011 2012
##    33   49
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    17   21   17   18   13   20   12   11   17   13   25   19   17   29   29
## 2011 2012
##    22   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    11   18   14   11   11   18   11   11   15   11   22   17   14   26   22
## 2011 2012
##    21   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 = 20, df = 16, p-value = 0.2

```



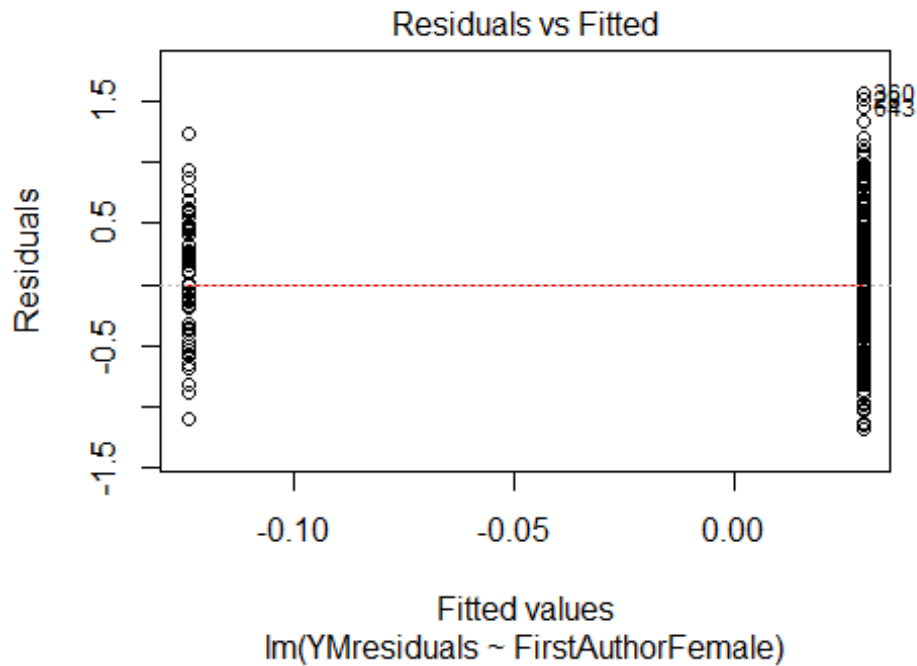
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 2.23606797749979"
## [1] "Male first author team size 2018 geometric mean: 2.00627388039457"

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

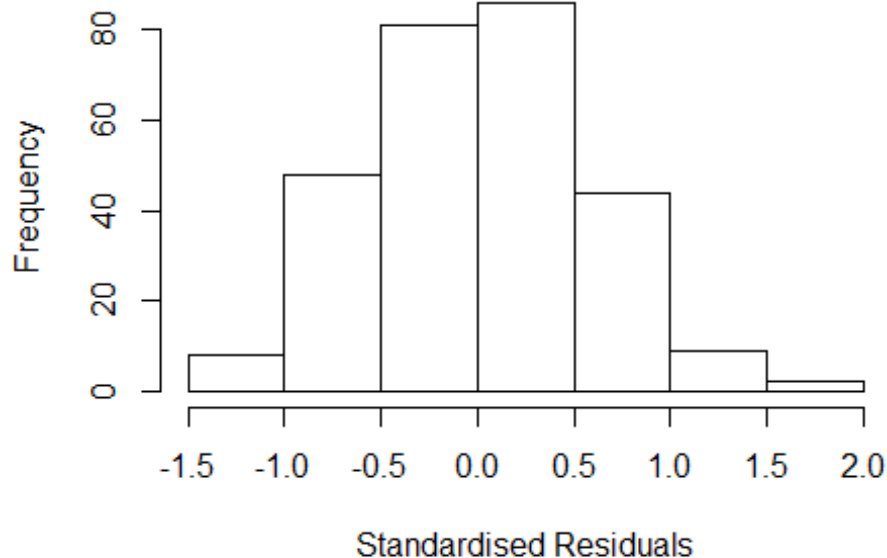
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.51486685936587"
## [1] "Male last author team size 2018 geometric mean: 1.91647392999983"

## 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 = 0.4
## 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 2.102  1         1.450
## LastAuthorFemale  2.577  1         1.605
## UniqueAuthors    2.930  4         1.144
## Year             3.644 16         1.041
```


Residuals from first and last author and team size



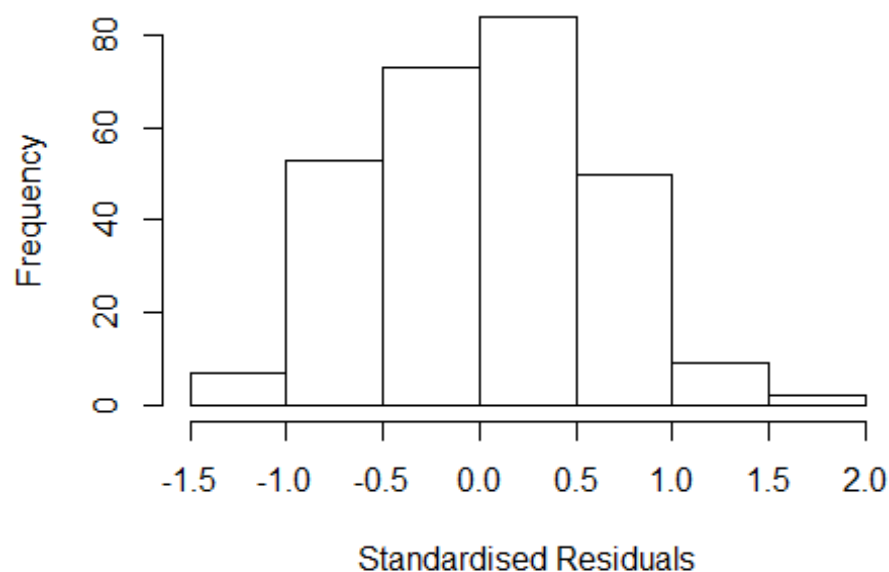
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3670 -0.4210 0.0131 0.3935 1.7932
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02350 0.34937 2.93 0.0037 **
## FirstAuthorFemale1 -0.06157 0.10579 -0.58 0.5611
## LastAuthorFemale1 -0.15542 0.10738 -1.45 0.1490
## UniqueAuthors2 0.08016 0.09271 0.86 0.3881
## UniqueAuthors3 0.15932 0.10772 1.48 0.1404
## UniqueAuthors4 0.18329 0.23912 0.77 0.4441
## UniqueAuthors5 0.08537 0.32290 0.26 0.7917
## Year1997 -0.17474 0.36910 -0.47 0.6363
## Year1998 -0.34065 0.39386 -0.86 0.3879
## Year1999 0.02269 0.39512 0.06 0.9542
```

```

## Year2000      0.32490      0.37266      0.87      0.3841
## Year2001     -0.38547      0.38793     -0.99      0.3213
## Year2002     -0.47401      0.39559     -1.20      0.2319
## Year2003     -0.34545      0.41584     -0.83      0.4069
## Year2004      0.04189      0.37504      0.11      0.9112
## Year2005     -0.19621      0.36172     -0.54      0.5880
## Year2006     -0.16375      0.36020     -0.45      0.6498
## Year2007      0.13765      0.36185      0.38      0.7040
## Year2008     -0.04990      0.38680     -0.13      0.8975
## Year2009     -0.08732      0.38329     -0.23      0.8200
## Year2010      0.00113      0.35431      0.00      0.9975
## Year2011     -0.25836      0.36141     -0.71      0.4753
## Year2012     -0.37716      0.35915     -1.05      0.2946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0347
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 261 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.345  0.879  0.951  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      3.60e-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 1.962 1      1.401
## LastAuthorFemale  2.165 1      1.471
## Year              1.432 16      1.011

```

Residuals from first and last author



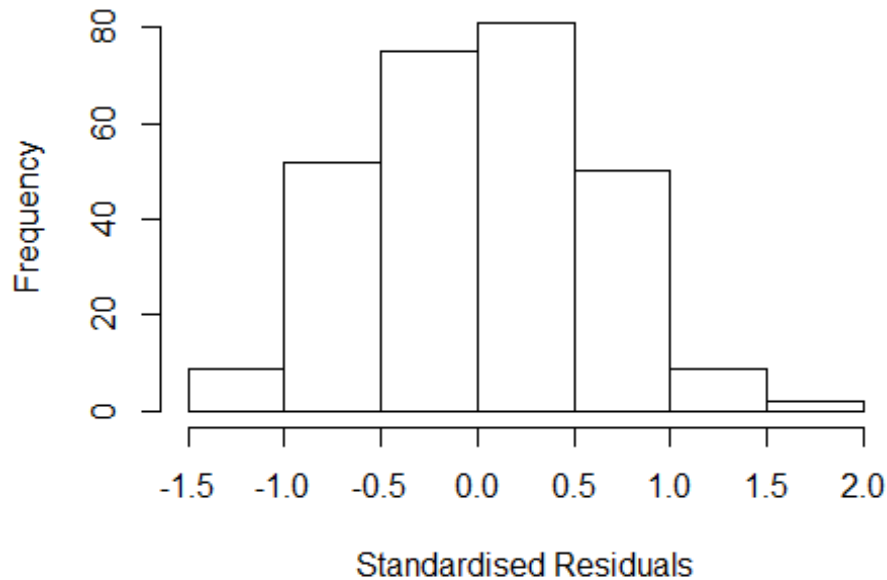
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3001 -0.4187 0.0302 0.4319 1.7579
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09685 0.35837 3.06 0.0024 **
## FirstAuthorFemale1 -0.06952 0.10707 -0.65 0.5167
## LastAuthorFemale1 -0.11060 0.10093 -1.10 0.2742
## Year1997 -0.22468 0.37845 -0.59 0.5532
## Year1998 -0.37198 0.39886 -0.93 0.3519
## Year1999 -0.01616 0.40419 -0.04 0.9681
## Year2000 0.27273 0.38194 0.71 0.4758
## Year2001 -0.42668 0.39438 -1.08 0.2803
## Year2002 -0.52943 0.40623 -1.30 0.1936
## Year2003 -0.34813 0.41528 -0.84 0.4026
## Year2004 -0.00102 0.38388 0.00 0.9979
## Year2005 -0.25175 0.37174 -0.68 0.4989
```

```

## Year2006      -0.18638    0.37276   -0.50    0.6175
## Year2007      0.08515    0.37132    0.23    0.8188
## Year2008     -0.08280    0.39766   -0.21    0.8352
## Year2009     -0.10269    0.39769   -0.26    0.7964
## Year2010     -0.02635    0.36615   -0.07    0.9427
## Year2011     -0.28509    0.37136   -0.77    0.4434
## Year2012     -0.37895    0.36922   -1.03    0.3057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.603
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0392
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.376  0.889  0.947  0.918  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      3.60e-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 1.092 1      1.045
## Year      1.092 16      1.003

```

Residuals from first author



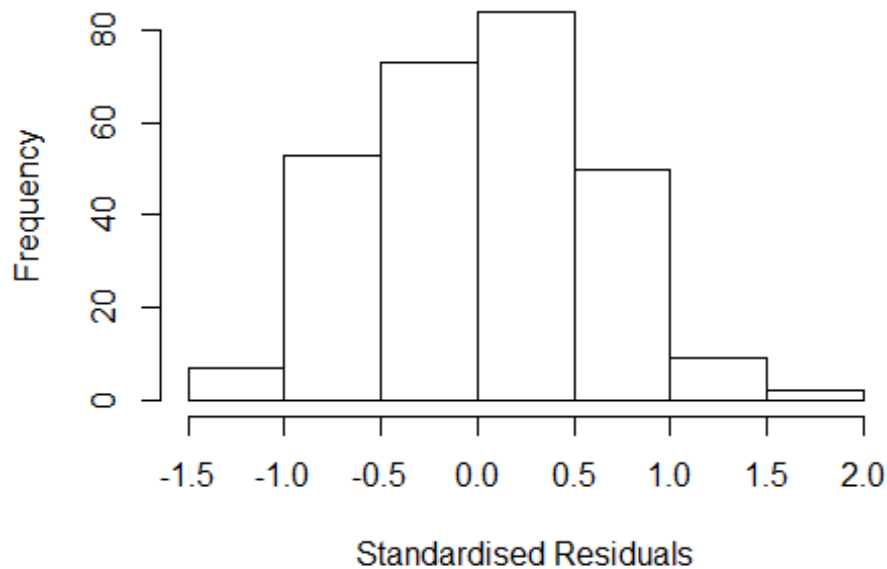
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2014 -0.4227 0.0287 0.4435 1.6467
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06579 0.35292 3.02 0.0028 **
## FirstAuthorFemale1 -0.12980 0.08253 -1.57 0.1170
## Year1997 -0.20668 0.37520 -0.55 0.5822
## Year1998 -0.36362 0.39628 -0.92 0.3597
## Year1999 -0.02781 0.40461 -0.07 0.9453
## Year2000 0.26546 0.37555 0.71 0.4803
## Year2001 -0.41781 0.39293 -1.06 0.2886
## Year2002 -0.50945 0.40663 -1.25 0.2114
## Year2003 -0.31645 0.39834 -0.79 0.4277
## Year2004 -0.00742 0.38657 -0.02 0.9847
## Year2005 -0.25389 0.36994 -0.69 0.4931
## Year2006 -0.17045 0.36937 -0.46 0.6448
```

```

## Year2007          0.09657    0.37133    0.26    0.7950
## Year2008          -0.06059    0.39676   -0.15    0.8787
## Year2009          -0.07297    0.39496   -0.18    0.8536
## Year2010          -0.00245    0.36262   -0.01    0.9946
## Year2011          -0.25971    0.36660   -0.71    0.4793
## Year2012          -0.36274    0.36692   -0.99    0.3238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0978, Adjusted R-squared:  0.0388
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.418  0.885  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      3.60e-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 1.233 1      1.110
## Year              1.233 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
## -1.3636 -0.3954 0.0276 0.4183 1.8265
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.092056 0.362329 3.01 0.0028 **
## LastAuthorFemale1 -0.144781 0.077317 -1.87 0.0623 .
## Year1997 -0.228084 0.380449 -0.60 0.5494
## Year1998 -0.373599 0.401188 -0.93 0.3526
## Year1999 -0.000549 0.405090 0.00 0.9989
## Year2000 0.271494 0.385132 0.70 0.4815
## Year2001 -0.427749 0.399801 -1.07 0.2857
## Year2002 -0.535622 0.409788 -1.31 0.1923
## Year2003 -0.377768 0.420631 -0.90 0.3700
## Year2004 0.005728 0.386365 0.01 0.9882
## Year2005 -0.242418 0.376314 -0.64 0.5200
## Year2006 -0.179934 0.376498 -0.48 0.6331
```

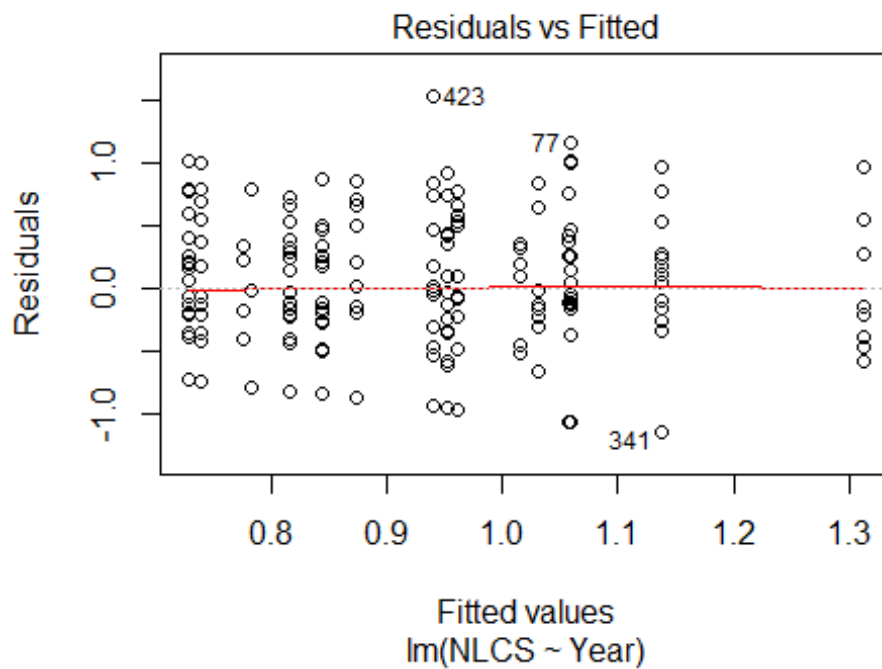
```

## Year2007      0.087908    0.373030    0.24    0.8139
## Year2008     -0.091601    0.401788   -0.23    0.8198
## Year2009     -0.104887    0.402040   -0.26    0.7944
## Year2010     -0.026265    0.369105   -0.07    0.9433
## Year2011     -0.288079    0.375047   -0.77    0.4431
## Year2012     -0.382009    0.371906   -1.03    0.3053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.103, Adjusted R-squared:  0.0446
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.310  0.880  0.943  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      3.60e-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: 278"
## [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
##   14   25   10   10   16   16   10   15   15   11   19   25   24   24   22
## 2011 2012
##   26   21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   14    3    6   10   11    4    6    8    8   15   23   15   17   14
## 2011 2012

```



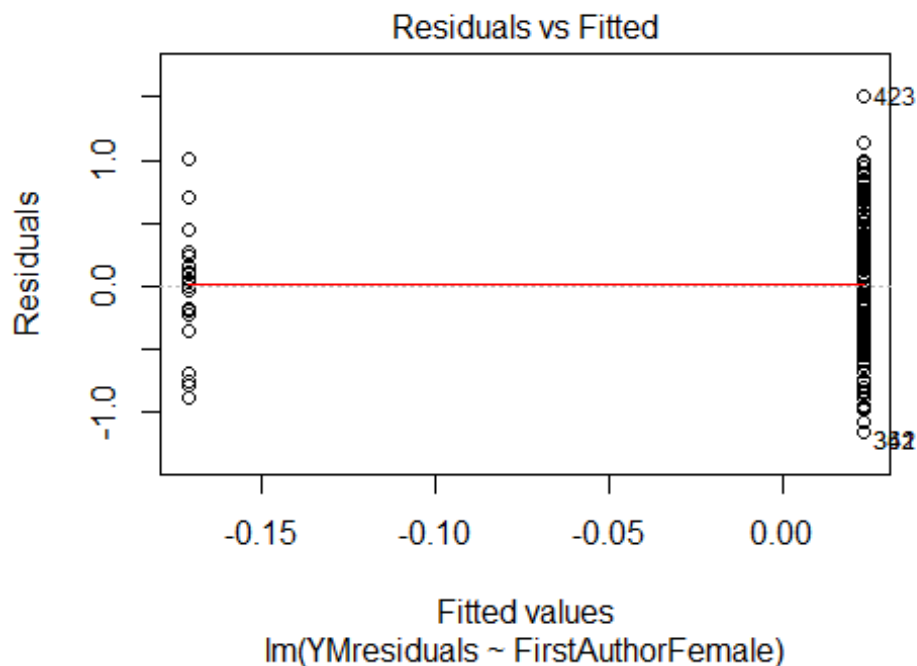
```
## 19 14
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 12 3 5 8 11 4 6 8 7 14 22 14 12 13
## 2011 2012
## 16 14
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 12, df = 16, p-value = 0.8
```



```
##
## 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: 1"
## [1] "Male first author team size 2018 geometric mean: 1.23007550557797"
## 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 = 4.5, p-value = 0.7
## 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: 1.13431252219546"

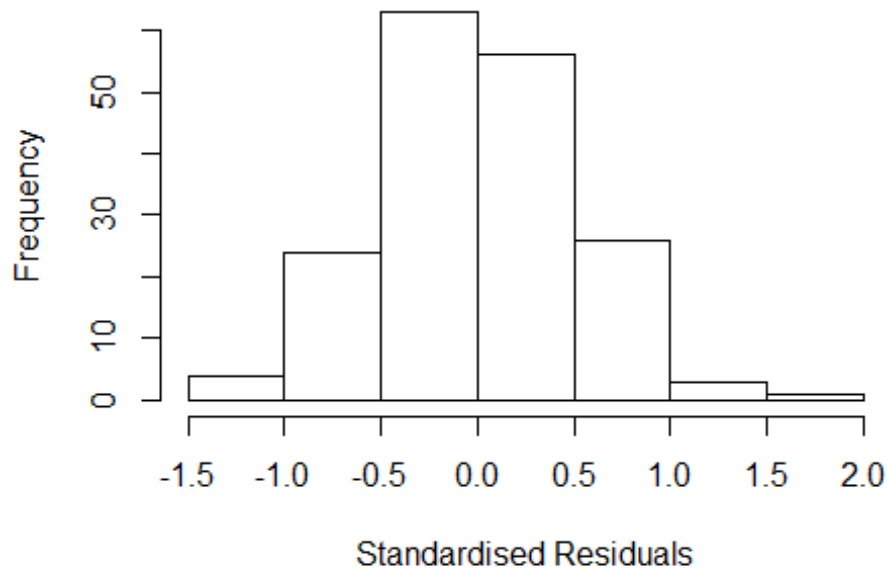
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16, p-value = 0.3
## 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	2.610	1	1.616
LastAuthorFemale	2.229	1	1.493
UniqueAuthors	2.848	3	1.191
Year	5.260	16	1.053

Residuals from first and last author and team size



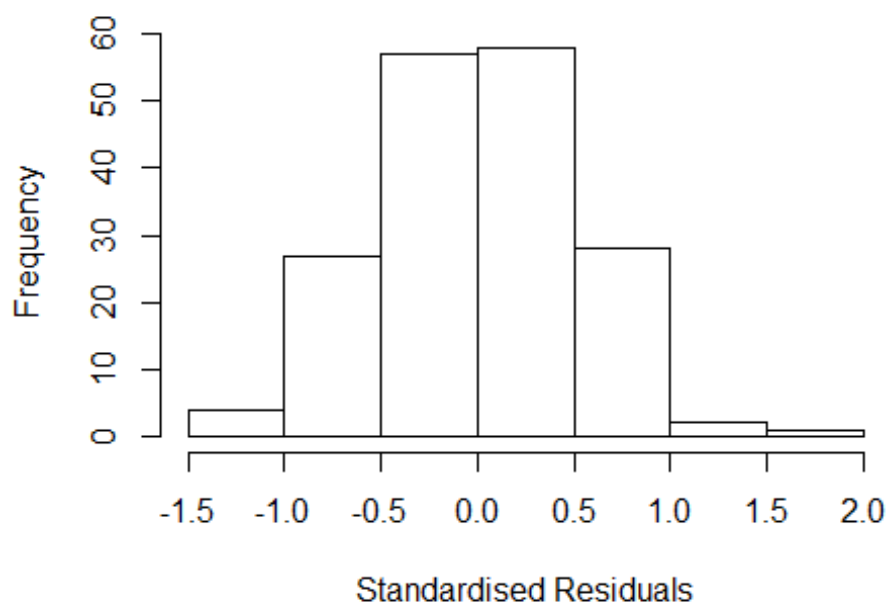
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.068 -0.357 -0.015 0.346 1.677
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.207 0.236 5.12 9.1e-07 ***
## FirstAuthorFemale1 -0.369 0.151 -2.44 0.016 *
## LastAuthorFemale1 0.181 0.176 1.03 0.306
## UniqueAuthors2 0.253 0.097 2.61 0.010 *
## UniqueAuthors3 0.080 0.135 0.59 0.553
## UniqueAuthors4 0.823 0.164 5.03 1.3e-06 ***
## Year1997 -0.365 0.294 -1.24 0.216
## Year1998 -0.577 0.548 -1.05 0.294
## Year1999 -0.139 0.417 -0.33 0.741
## Year2000 -0.114 0.292 -0.39 0.697
```

```

## Year2001          -0.405      0.289   -1.40    0.163
## Year2002          -0.374      0.256   -1.46    0.146
## Year2003          -0.271      0.257   -1.05    0.294
## Year2004          -0.201      0.270   -0.74    0.457
## Year2005          -0.163      0.322   -0.51    0.614
## Year2006          -0.390      0.265   -1.47    0.143
## Year2007          -0.481      0.254   -1.90    0.060 .
## Year2008          -0.562      0.266   -2.11    0.036 *
## Year2009          -0.383      0.270   -1.42    0.157
## Year2010          -0.189      0.271   -0.70    0.485
## Year2011          -0.554      0.257   -2.15    0.033 *
## Year2012          -0.415      0.262   -1.58    0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.165, Adjusted R-squared:  0.0517
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 162 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.353  0.884  0.962  0.922  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07          5.65e-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 1.987 1      1.410
## LastAuthorFemale  1.688 1      1.299
## Year              2.037 16      1.022

```

Residuals from first and last author



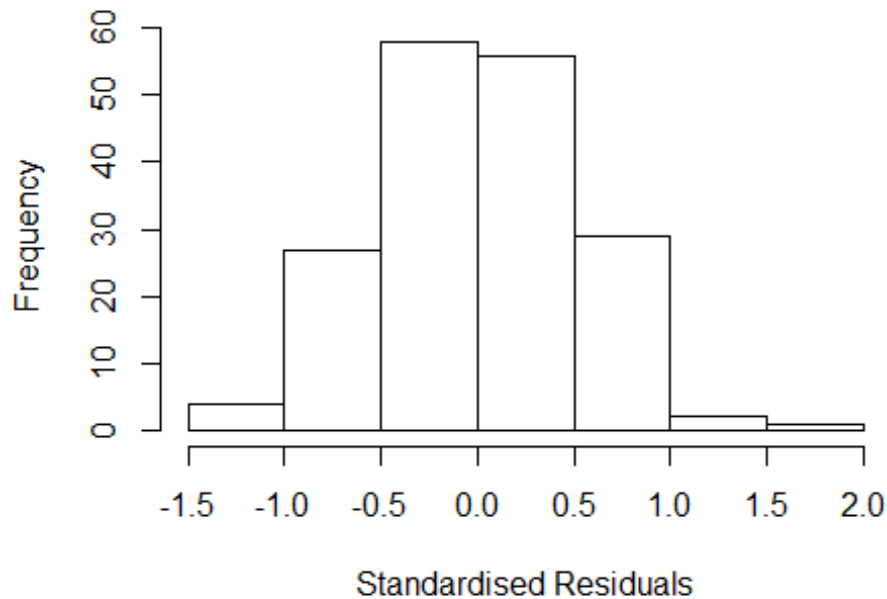
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2721 -0.3428 0.0116 0.3654 1.5555
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3228 0.2136 6.19 4.9e-09 ***
## FirstAuthorFemale1 -0.2634 0.1425 -1.85 0.0664 .
## LastAuthorFemale1 0.0941 0.1737 0.54 0.5887
## Year1997 -0.4191 0.2829 -1.48 0.1405
## Year1998 -0.5415 0.4980 -1.09 0.2785
## Year1999 -0.1756 0.4079 -0.43 0.6675
## Year2000 -0.1988 0.2693 -0.74 0.4616
## Year2001 -0.4194 0.3012 -1.39 0.1658
## Year2002 -0.4367 0.2335 -1.87 0.0633 .
## Year2003 -0.3004 0.2629 -1.14 0.2549
## Year2004 -0.2558 0.2654 -0.96 0.3368
## Year2005 -0.0507 0.2874 -0.18 0.8603
```

```

## Year2006          -0.4718      0.2485    -1.90    0.0595 .
## Year2007          -0.5624      0.2361    -2.38    0.0184 *
## Year2008          -0.6588      0.2461    -2.68    0.0082 **
## Year2009          -0.3761      0.2698    -1.39    0.1652
## Year2010          -0.2178      0.2654    -0.82    0.4130
## Year2011          -0.5208      0.2433    -2.14    0.0339 *
## Year2012          -0.4092      0.2787    -1.47    0.1439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.103, Adjusted R-squared:  0.000412
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 162 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.462  0.873   0.959   0.923   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      5.65e-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 1.538 1      1.240
## Year              1.538 16      1.014

```

Residuals from first author

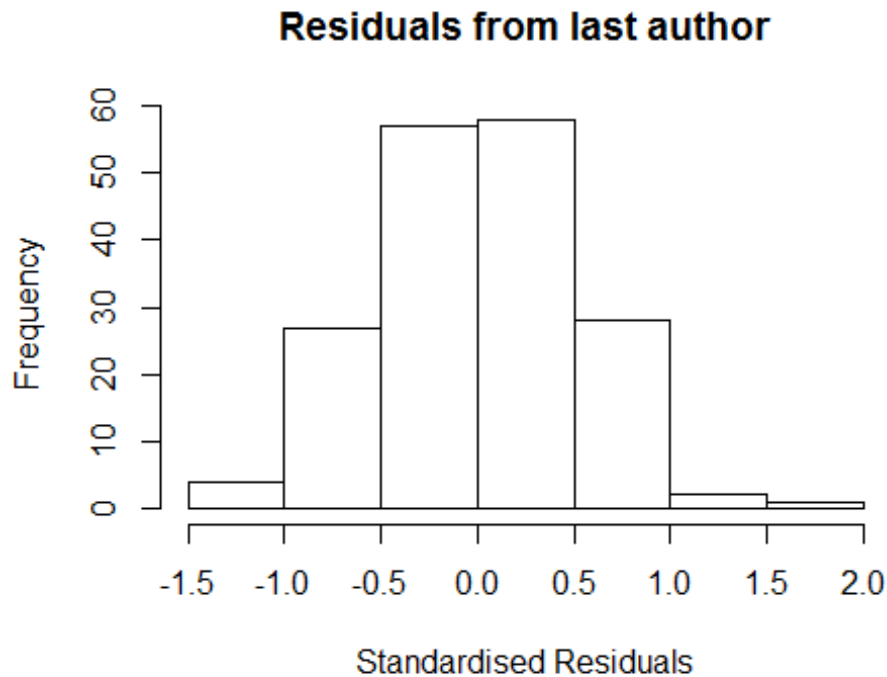


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27314 -0.34819 -0.00621 0.33439 1.54916
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3465 0.2128 6.33 2.4e-09 ***
## FirstAuthorFemale1 -0.2106 0.1284 -1.64 0.1029
## Year1997 -0.4352 0.2858 -1.52 0.1298
## Year1998 -0.5653 0.4990 -1.13 0.2590
## Year1999 -0.1990 0.4079 -0.49 0.6263
## Year2000 -0.2340 0.2695 -0.87 0.3866
## Year2001 -0.4397 0.2995 -1.47 0.1440
## Year2002 -0.4642 0.2274 -2.04 0.0429 *
## Year2003 -0.3240 0.2624 -1.23 0.2187
## Year2004 -0.2733 0.2692 -1.02 0.3115
## Year2005 -0.0733 0.2869 -0.26 0.7986
## Year2006 -0.4956 0.2479 -2.00 0.0473 *
```

```

## Year2007          -0.5896      0.2329   -2.53   0.0123 *
## Year2008          -0.6710      0.2523   -2.66   0.0086 **
## Year2009          -0.4095      0.2607   -1.57   0.1181
## Year2010          -0.2411      0.2649   -0.91   0.3641
## Year2011          -0.5359      0.2473   -2.17   0.0317 *
## Year2012          -0.4266      0.2786   -1.53   0.1277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.583
## Multiple R-squared:  0.102, Adjusted R-squared:  0.00593
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 162 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.459  0.873  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      5.65e-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 1.315 1          1.147
## Year            1.315 16          1.009

```

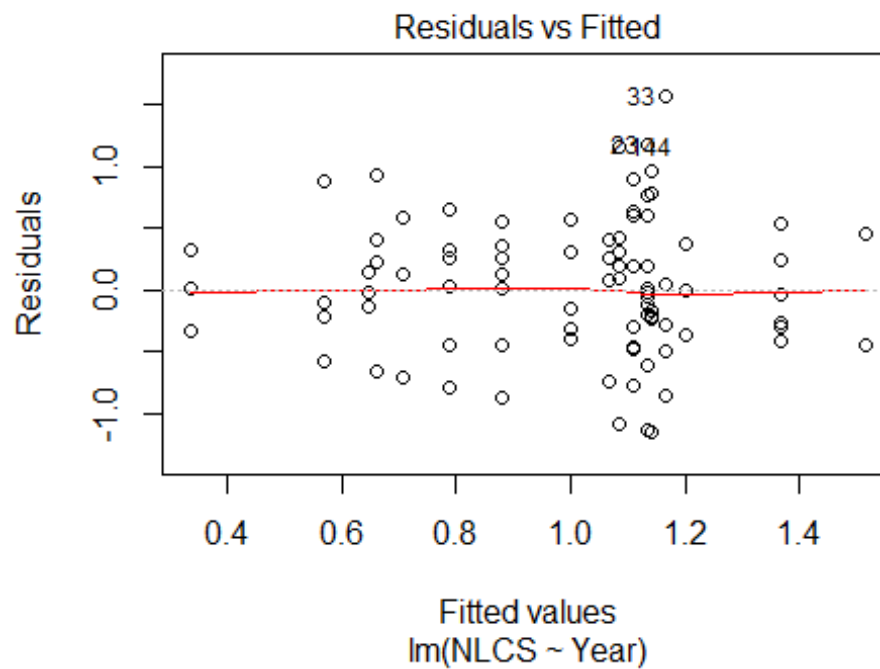
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27325 -0.34549 -0.00933 0.33322 1.57034
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3228 0.2210 5.98 1.4e-08 ***
## LastAuthorFemale1 -0.0929 0.1608 -0.58 0.564
## Year1997 -0.4323 0.2922 -1.48 0.141
## Year1998 -0.5416 0.5027 -1.08 0.283
## Year1999 -0.1753 0.4123 -0.43 0.671
## Year2000 -0.2529 0.2894 -0.87 0.384
## Year2001 -0.4245 0.3121 -1.36 0.176
## Year2002 -0.5232 0.2529 -2.07 0.040 *
## Year2003 -0.3003 0.2692 -1.12 0.266
## Year2004 -0.2926 0.2782 -1.05 0.295
## Year2005 -0.0495 0.2930 -0.17 0.866
## Year2006 -0.4719 0.2551 -1.85 0.066 .
```

```

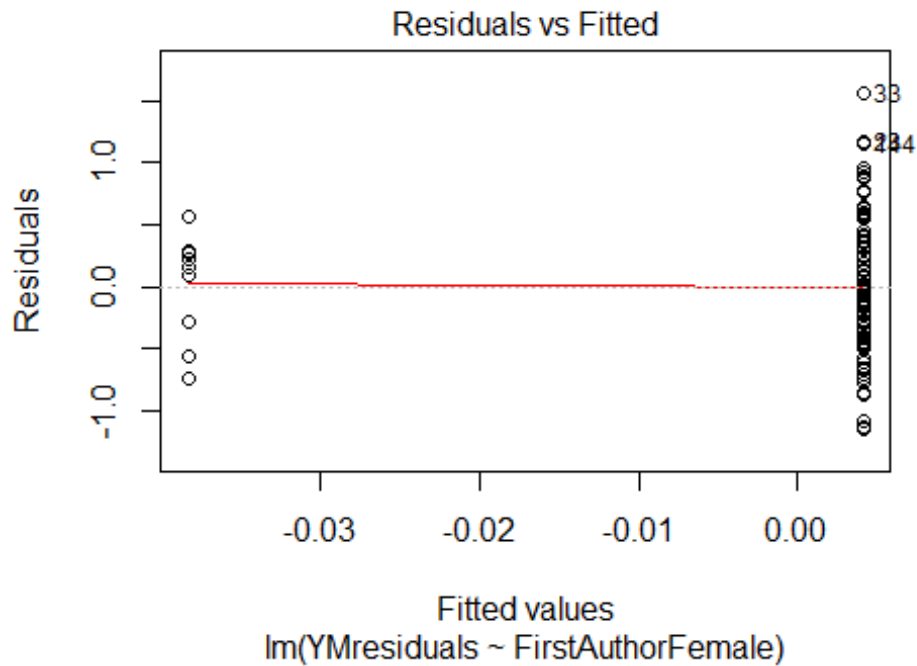
## Year2007          -0.5953      0.2411   -2.47    0.015 *
## Year2008          -0.6494      0.2519   -2.58    0.011 *
## Year2009          -0.4268      0.2679   -1.59    0.113
## Year2010          -0.2173      0.2715   -0.80    0.425
## Year2011          -0.5110      0.2460   -2.08    0.039 *
## Year2012          -0.4241      0.2869   -1.48    0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.0916, Adjusted R-squared:  -0.00557
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 163 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.447  0.879  0.959  0.921  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      5.65e-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: 177"
## [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
##    4   15    8    7   10   10    6    3    3    3    7    8    8   12    9
## 2011 2012
##    7    7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3   12    5    3    5    4    4    3    3    2    6    7    7    9    7
## 2011 2012

```

```
##      6      7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      3     11     5     3     5     4     4     3     3     2     5     5     7     9     7
## 2011 2012
##      6      7
## [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
```

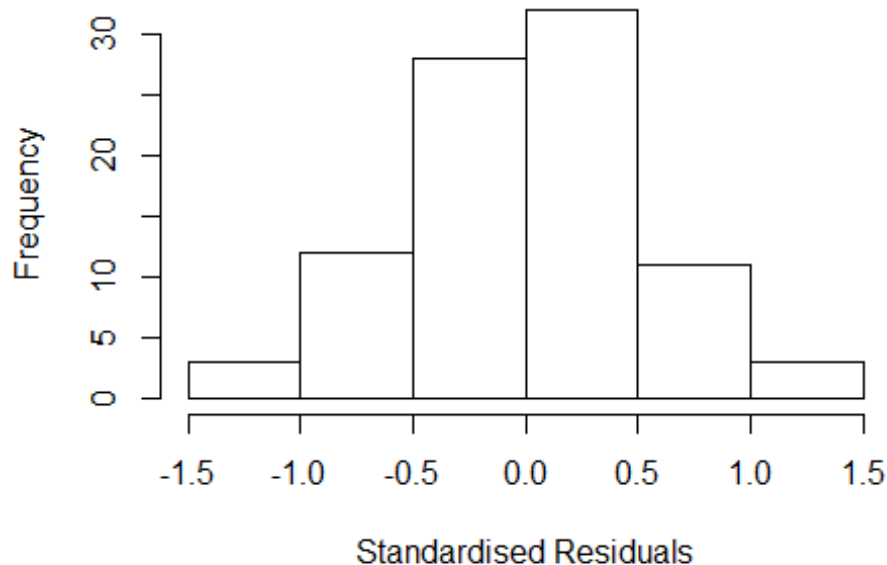


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.84, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: 1"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale  37.672  1         6.138
## LastAuthorFemale   2.872  1         1.695
## UniqueAuthors    1106.317  3         3.216
## Year              2952.737 16         1.284
```

Residuals from first and last author and team size



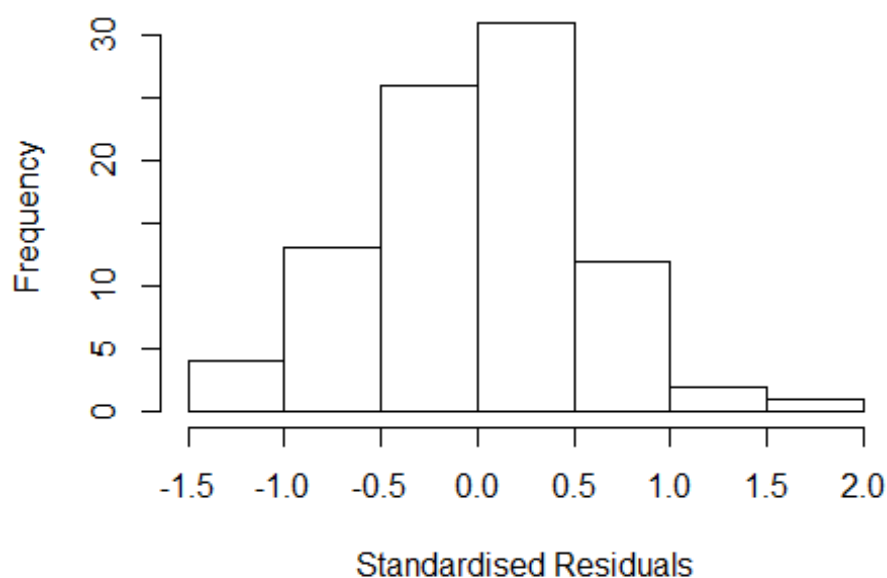
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1947 -0.3146 0.0435 0.3651 1.2013
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.33375 0.15818 2.11 0.03860 *
## FirstAuthorFemale1 0.37067 0.19764 1.88 0.06509 .
## LastAuthorFemale1 -0.63375 0.20325 -3.12 0.00268 **
## UniqueAuthors2 0.45802 0.18722 2.45 0.01706 *
## UniqueAuthors3 -0.69960 0.17443 -4.01 0.00015 ***
## UniqueAuthors4 -0.28378 0.42729 -0.66 0.50889
## Year1997 0.76997 0.26803 2.87 0.00545 **
## Year1998 0.80036 0.40731 1.96 0.05356 .
## Year1999 0.71573 0.19404 3.69 0.00045 ***
## Year2000 0.71682 0.23518 3.05 0.00330 **
```

```

## Year2001      0.29408      0.21240      1.38      0.17079
## Year2002      0.75552      0.27991      2.70      0.00879 **
## Year2003      0.32353      0.36158      0.89      0.37412
## Year2004      0.00262      0.29370      0.01      0.99291
## Year2005      0.95324      0.66149      1.44      0.15423
## Year2006      0.65540      0.24177      2.71      0.00852 **
## Year2007      0.86884      0.20453      4.25      6.8e-05 ***
## Year2008      0.48995      0.25208      1.94      0.05615 .
## Year2009      0.78326      0.28128      2.78      0.00696 **
## Year2010      0.86098      0.40140      2.14      0.03559 *
## Year2011      0.59457      0.30400      1.96      0.05466 .
## Year2012      0.37015      0.33629      1.10      0.27497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.595
## Multiple R-squared:  0.277, Adjusted R-squared:  0.0505
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 78 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.663  0.887  0.958  0.924  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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-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 3.910 1      1.977
## LastAuthorFemale 3.130 1      1.769
## Year      6.836 16      1.062

```

Residuals from first and last author



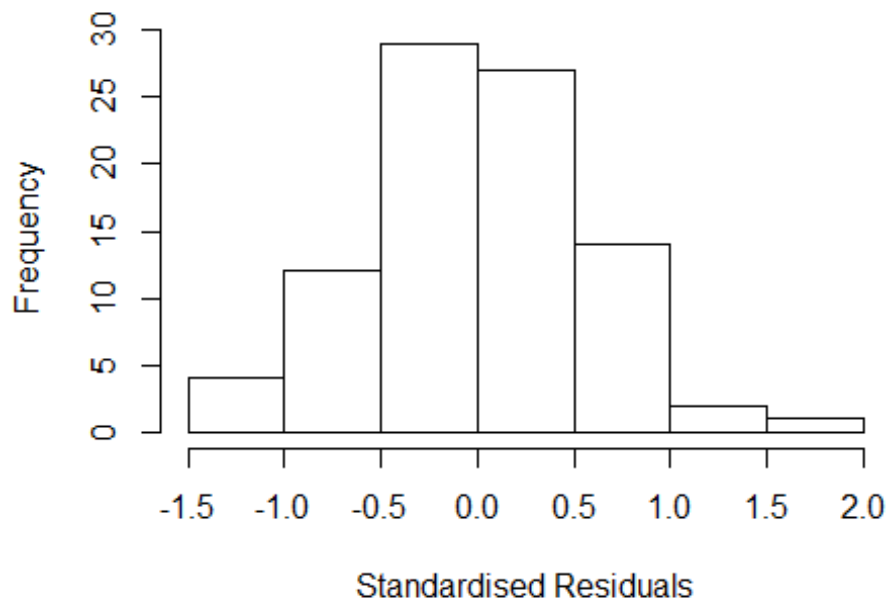
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18089 -0.37191  0.00846  0.33505  1.90896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.334      0.158    2.11  0.0384 *
## FirstAuthorFemale1 0.389      0.211    1.85  0.0687 .
## LastAuthorFemale1 -0.592      0.243   -2.43  0.0176 *
## Year1997          0.815      0.256    3.19  0.0021 **
## Year1998          0.489      0.360    1.36  0.1780
## Year1999          0.868      0.238    3.65  0.0005 ***
## Year2000          0.704      0.236    2.98  0.0039 **
## Year2001          0.199      0.328    0.61  0.5465
## Year2002          0.755      0.280    2.70  0.0087 **
## Year2003          0.472      0.432    1.09  0.2779
## Year2004          0.314      0.171    1.83  0.0710 .
## Year2005          1.182      0.387    3.06  0.0032 **
```

```

## Year2006          0.495      0.280      1.77      0.0818 .
## Year2007          0.960      0.202      4.76      1e-05 ***
## Year2008          0.566      0.241      2.35      0.0217 *
## Year2009          0.842      0.277      3.03      0.0034 **
## Year2010          0.847      0.398      2.13      0.0370 *
## Year2011          0.815      0.371      2.20      0.0312 *
## Year2012          0.360      0.331      1.09      0.2808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.597
## Multiple R-squared:  0.196, Adjusted R-squared:  -0.0105
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 79 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
## 0.285 0.891 0.959 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      1.12e-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 1.933 1      1.390
## Year              1.933 16      1.021

```


Residuals from first author



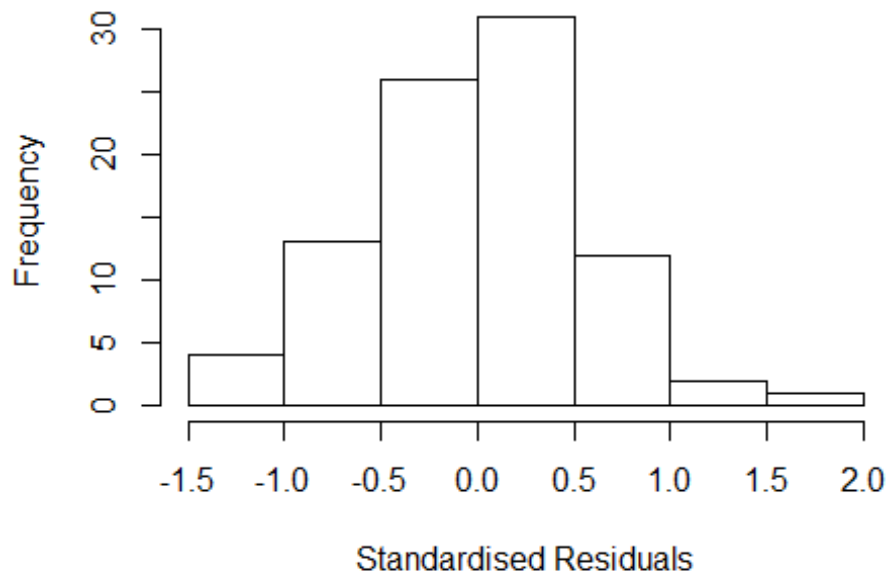
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14798 -0.40672 -0.00258 0.35953 1.74925
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3337 0.1579 2.11 0.03804 *
## FirstAuthorFemale1 -0.0568 0.2105 -0.27 0.78804
## Year1997 0.8010 0.2586 3.10 0.00279 **
## Year1998 0.6490 0.4078 1.59 0.11594
## Year1999 0.8679 0.2374 3.66 0.00049 ***
## Year2000 0.6730 0.2401 2.80 0.00651 **
## Year2001 0.2012 0.3264 0.62 0.53943
## Year2002 0.7539 0.2790 2.70 0.00862 **
## Year2003 0.4101 0.4132 0.99 0.32429
## Year2004 0.3139 0.1709 1.84 0.07051 .
## Year2005 1.1823 0.3845 3.08 0.00299 **
## Year2006 0.4706 0.2960 1.59 0.11627
```

```

## Year2007          1.0439      0.2298      4.54  2.2e-05 ***
## Year2008          0.5648      0.2403      2.35  0.02157 *
## Year2009          0.7603      0.2650      2.87  0.00542 **
## Year2010          0.8098      0.3875      2.09  0.04024 *
## Year2011          0.8143      0.3671      2.22  0.02973 *
## Year2012          0.3337      0.3204      1.04  0.30113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.168, Adjusted R-squared:  -0.0314
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 80 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.400  0.896  0.959  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      1.12e-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 1.782 1      1.335
## Year              1.782 16      1.018

```

Residuals from last author



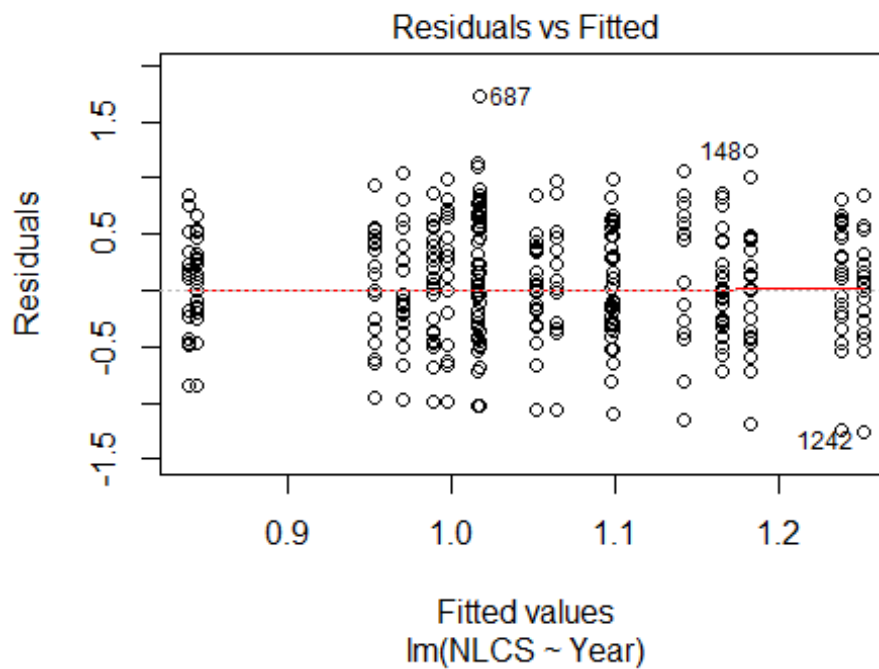
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18922 -0.36330 -0.00759 0.36493 1.76800
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.334 0.158 2.11 0.03806 *
## LastAuthorFemale1 -0.267 0.187 -1.43 0.15756
## Year1997 0.823 0.253 3.25 0.00177 **
## Year1998 0.630 0.375 1.68 0.09687 .
## Year1999 0.868 0.237 3.66 0.00049 ***
## Year2000 0.718 0.233 3.08 0.00292 **
## Year2001 0.201 0.326 0.62 0.53978
## Year2002 0.754 0.279 2.70 0.00861 **
## Year2003 0.496 0.432 1.15 0.25433
## Year2004 0.314 0.171 1.84 0.07054 .
## Year2005 1.182 0.385 3.07 0.00300 **
## Year2006 0.506 0.272 1.86 0.06757 .
```

```

## Year2007          1.034      0.226      4.58      2e-05 ***
## Year2008          0.565      0.240      2.35      0.02155 *
## Year2009          0.797      0.265      3.00      0.00369 **
## Year2010          0.855      0.395      2.17      0.03364 *
## Year2011          0.814      0.367      2.22      0.02983 *
## Year2012          0.370      0.332      1.12      0.26796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.18,   Adjusted R-squared:  -0.0159
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 80 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.388  0.898  0.956  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      1.12e-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: 89"
## [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
##   77   58   73   70   63   78   53   97   65   48   78   53   52   44   40
## 2011 2012
##   35   45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   20   19   23   27   20   15   32   21   20   34   21   26   24   20
## 2011 2012

```

```
## 15 28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 18 13 22 23 18 12 28 18 16 27 19 23 21 19
## 2011 2012
## 15 27
## [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 = 0.56, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 1.81712059283214"
## [1] "Male first author team size 2018 geometric mean: 1.82230011578833"

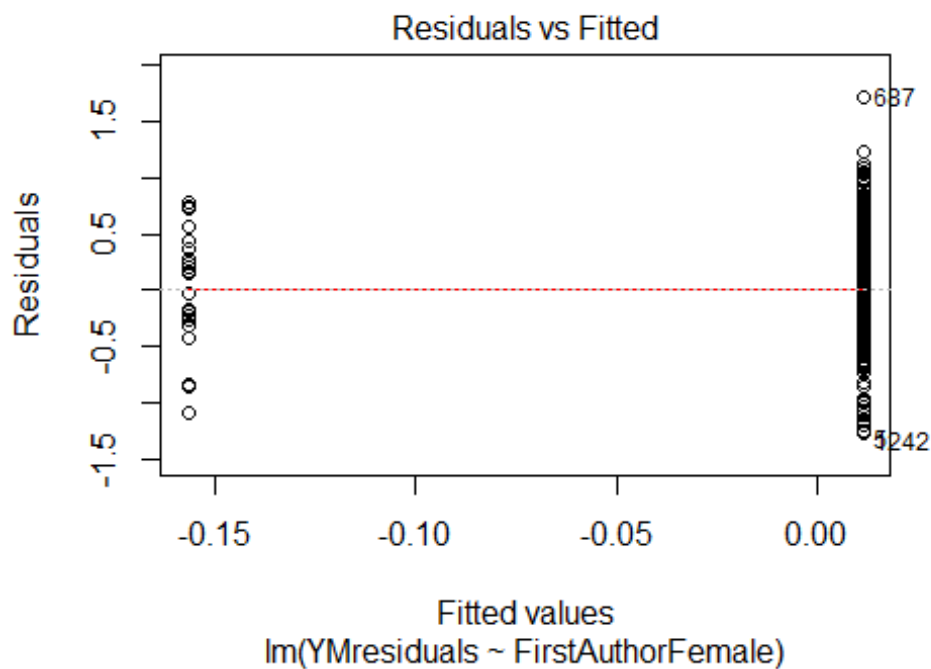
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 39, p-value = 1
## 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: 1.8022480601541"

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

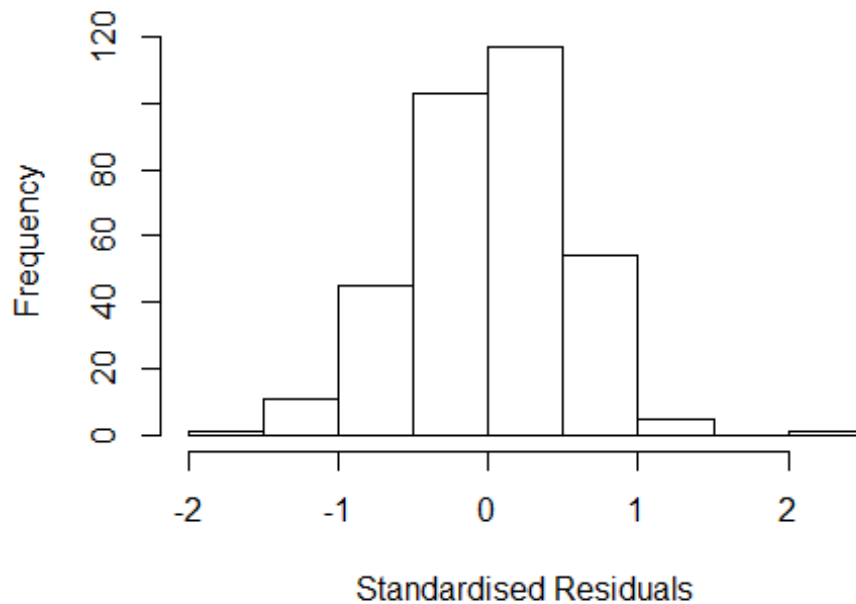
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'UniqueAuthors5'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```



##		GVIF	Df	GVIF^(1/(2*Df))
##	FirstAuthorFemale	1.753	1	1.324
##	LastAuthorFemale	1.563	1	1.250
##	UniqueAuthors	167.550	4	1.897
##	Year	217.640	16	1.183

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
## -1.6680 -0.3497 0.0214 0.3548 2.4980
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2034 0.1475 8.16 8.3e-15 ***
## FirstAuthorFemale1 -0.1587 0.1168 -1.36 0.17519
## LastAuthorFemale1 -0.0898 0.1111 -0.81 0.41935
## UniqueAuthors2 0.0094 0.0740 0.13 0.89894
```

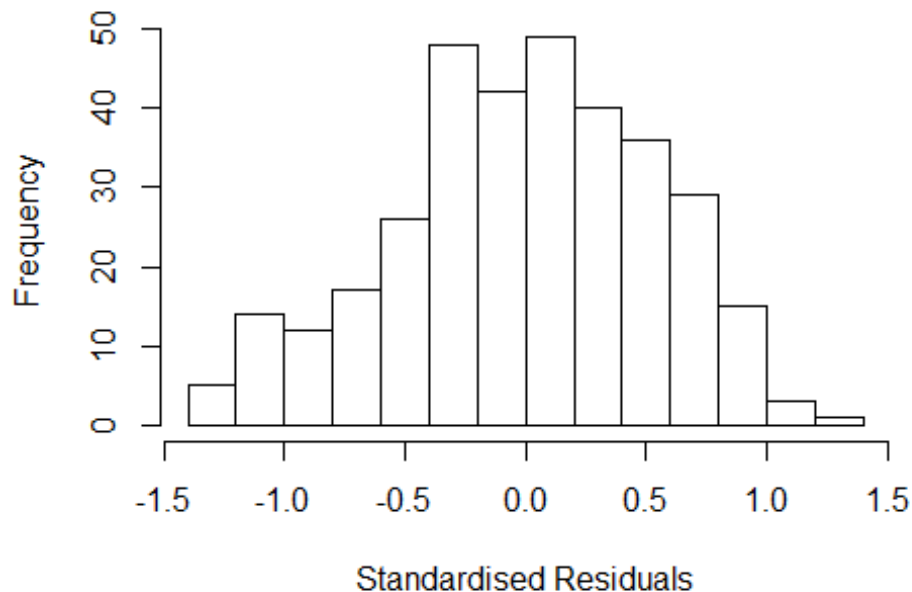
```

## UniqueAuthors3      0.3217      0.0952      3.38  0.00082 ***
## UniqueAuthors4      0.2198      0.1460      1.51  0.13329
## UniqueAuthors5     -1.3463      0.2441     -5.51  7.3e-08 ***
## Year1997            -0.0884      0.1891     -0.47  0.64060
## Year1998            -0.2542      0.1884     -1.35  0.17820
## Year1999            -0.1432      0.1695     -0.84  0.39889
## Year2000            -0.1992      0.1596     -1.25  0.21290
## Year2001             0.1429      0.2839      0.50  0.61526
## Year2002            -0.3232      0.2244     -1.44  0.15075
## Year2003            -0.2612      0.2279     -1.15  0.25255
## Year2004            -0.2480      0.2207     -1.12  0.26188
## Year2005            -0.3573      0.1747     -2.04  0.04172 *
## Year2006            -0.2507      0.1811     -1.38  0.16723
## Year2007            -0.2626      0.1913     -1.37  0.17080
## Year2008            -0.2641      0.1628     -1.62  0.10575
## Year2009            -0.3609      0.1685     -2.14  0.03295 *
## Year2010             0.0356      0.1663      0.21  0.83050
## Year2011            -0.4145      0.1825     -2.27  0.02377 *
## Year2012            -0.1708      0.1592     -1.07  0.28412
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.121, Adjusted R-squared:  0.0592
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 119 is an outlier with |weight| = 0 ( < 0.0003);
## 28 weights are ~= 1. The remaining 308 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.284  0.868  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      2.97e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))

```


## FirstAuthorFemale	1.379	1	1.174
## LastAuthorFemale	1.280	1	1.131
## Year	1.442	16	1.012

Residuals from first and last author



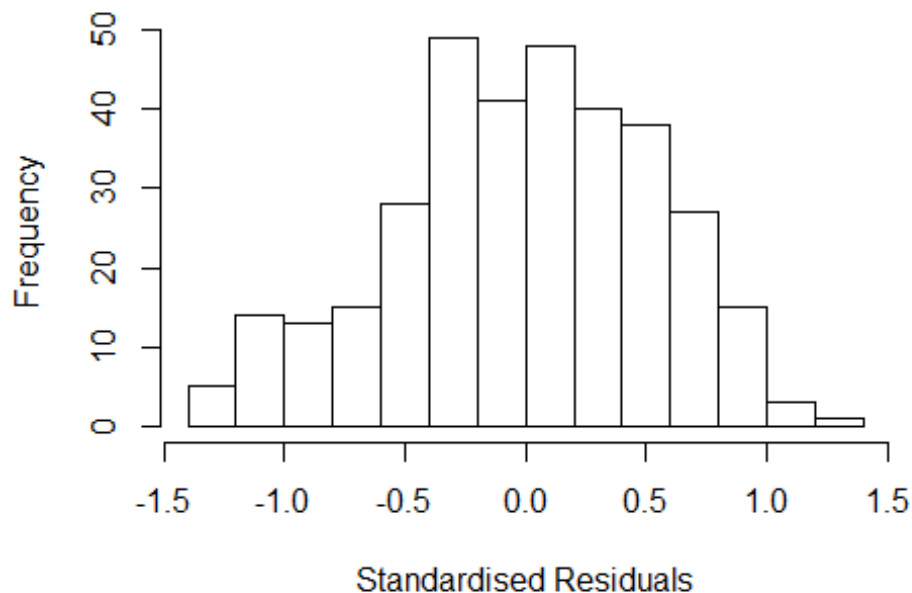
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2843 -0.3784 0.0162 0.3966 1.2718
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2678 0.1448 8.75 <2e-16 ***
## FirstAuthorFemale1 -0.1146 0.1144 -1.00 0.317
## LastAuthorFemale1 -0.1483 0.1103 -1.34 0.180
## Year1997 -0.1126 0.1998 -0.56 0.573
## Year1998 -0.2909 0.2026 -1.44 0.152
## Year1999 -0.1260 0.1858 -0.68 0.498
## Year2000 -0.1361 0.1774 -0.77 0.444
## Year2001 0.0003 0.2835 0.00 0.999
```

```

## Year2002          -0.1846      0.2317    -0.80      0.426
## Year2003          -0.2411      0.2345    -1.03      0.305
## Year2004          -0.1982      0.2249    -0.88      0.379
## Year2005          -0.3651      0.1805    -2.02      0.044 *
## Year2006          -0.2757      0.1843    -1.50      0.136
## Year2007          -0.2795      0.2039    -1.37      0.171
## Year2008          -0.2633      0.1742    -1.51      0.132
## Year2009          -0.3821      0.1750    -2.18      0.030 *
## Year2010           0.0165      0.1733      0.10      0.924
## Year2011          -0.4222      0.1972    -2.14      0.033 *
## Year2012          -0.1769      0.1623    -1.09      0.277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.055, Adjusted R-squared:  0.00152
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 303 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.543  0.857   0.941   0.904   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      2.97e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.257 1          1.121
## Year              1.257 16          1.007

```

Residuals from first author



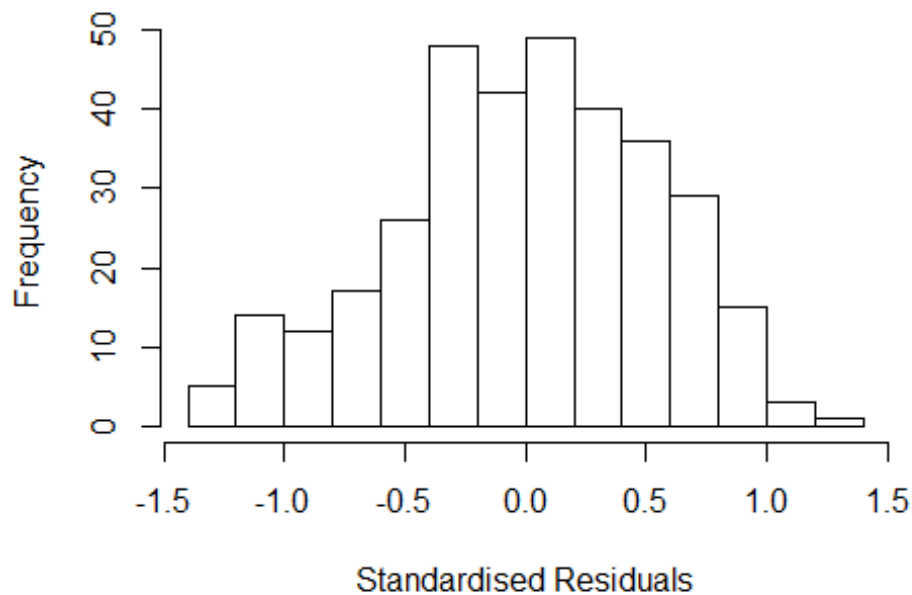
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2739 -0.3770 0.0229 0.3945 1.2727
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2543 0.1428 8.78 <2e-16 ***
## FirstAuthorFemale1 -0.1668 0.1102 -1.51 0.131
## Year1997 -0.1000 0.1993 -0.50 0.616
## Year1998 -0.2739 0.2031 -1.35 0.178
## Year1999 -0.1192 0.1860 -0.64 0.522
## Year2000 -0.1342 0.1736 -0.77 0.440
## Year2001 0.0196 0.2813 0.07 0.945
## Year2002 -0.1755 0.2335 -0.75 0.453
## Year2003 -0.2271 0.2337 -0.97 0.332
## Year2004 -0.1843 0.2238 -0.82 0.411
## Year2005 -0.3566 0.1759 -2.03 0.043 *
## Year2006 -0.2732 0.1839 -1.49 0.138
```

```

## Year2007          -0.2638      0.2012    -1.31     0.191
## Year2008          -0.2583      0.1746    -1.48     0.140
## Year2009          -0.3823      0.1733    -2.21     0.028 *
## Year2010           0.0148      0.1712     0.09     0.931
## Year2011          -0.4208      0.1969    -2.14     0.033 *
## Year2012          -0.1748      0.1614    -1.08     0.280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0515, Adjusted R-squared:  0.000958
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 303 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.546  0.858  0.938  0.902  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      2.97e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.159 1          1.077
## Year            1.159 16          1.005

```

Residuals from last author



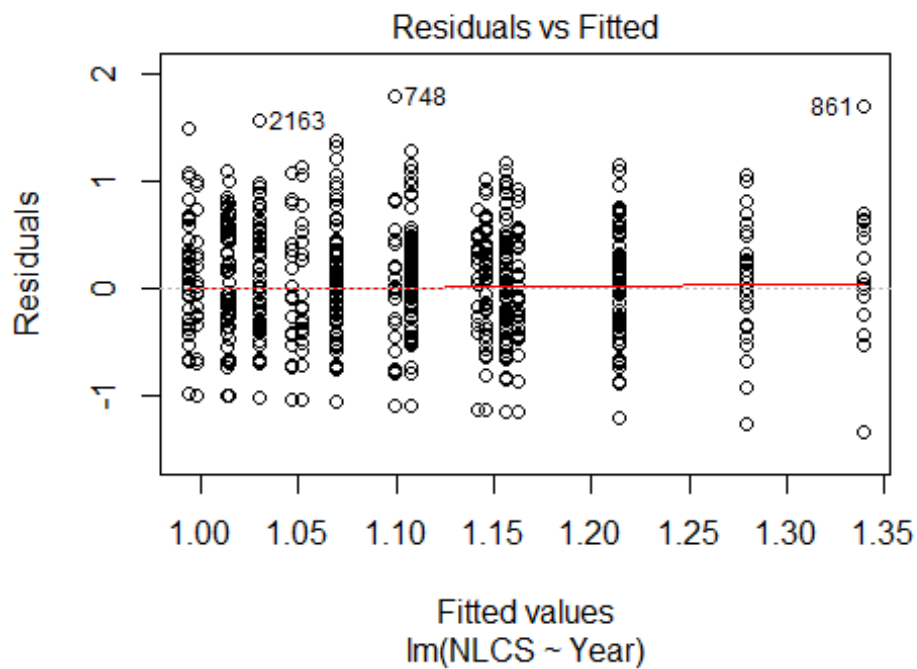
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.282 -0.372 0.010 0.409 1.291
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.258040 0.147849 8.51 6.9e-16 ***
## LastAuthorFemale1 -0.185730 0.105151 -1.77 0.078 .
## Year1997 -0.122169 0.202975 -0.60 0.548
## Year1998 -0.289789 0.201551 -1.44 0.151
## Year1999 -0.118619 0.186385 -0.64 0.525
## Year2000 -0.128726 0.181756 -0.71 0.479
## Year2001 0.000864 0.288739 0.00 0.998
## Year2002 -0.196927 0.232210 -0.85 0.397
## Year2003 -0.231189 0.236178 -0.98 0.328
## Year2004 -0.188304 0.226723 -0.83 0.407
## Year2005 -0.354306 0.183950 -1.93 0.055 .
## Year2006 -0.268109 0.186968 -1.43 0.153
```

```

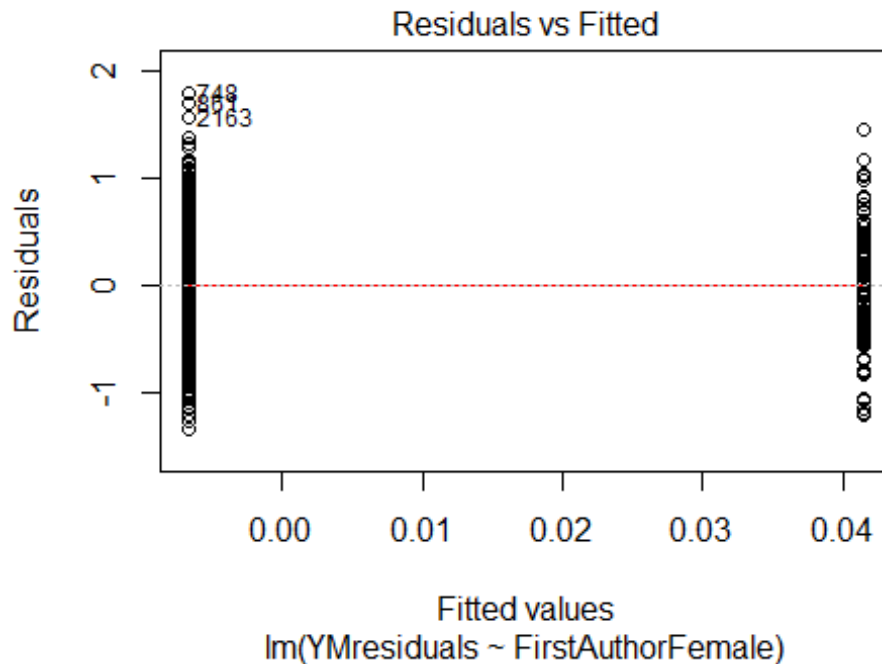
## Year2007          -0.287762    0.204024   -1.41    0.159
## Year2008          -0.269098    0.175800   -1.53    0.127
## Year2009          -0.374423    0.177623   -2.11    0.036 *
## Year2010           0.023736    0.176446    0.13    0.893
## Year2011          -0.409319    0.199233   -2.05    0.041 *
## Year2012          -0.164081    0.165108   -0.99    0.321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0529, Adjusted R-squared:  0.00241
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.538  0.860  0.935  0.902  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      2.97e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 337"
## [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
##   91   95   76   71  102  101   76   62   66   69  110  122  158  184  160
## 2011 2012
##  158  195
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   36   21   20   18   32   24   18   18   37   46   44   73   83   78
## 2011 2012

```

```
## 65 85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 33 21 16 12 29 22 15 16 32 39 37 63 66 60
## 2011 2012
## 53 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 = 10, df = 16, p-value = 0.9
```

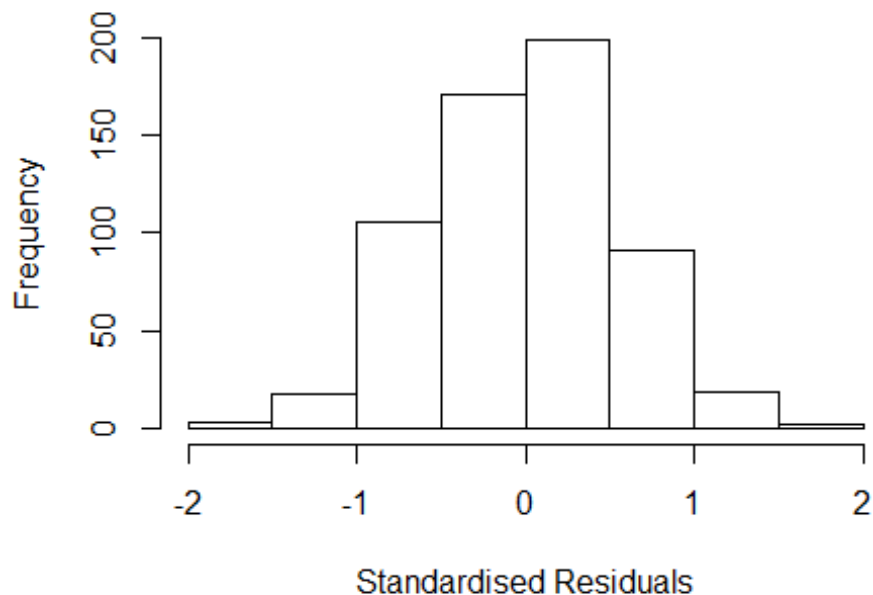


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.46, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 3.25239229643564"
## [1] "Male first author team size 2018 geometric mean: 2.49248506264797"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 590, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.01774902754934"
## [1] "Male last author team size 2018 geometric mean: 2.55187343538669"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, 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 1.385 1      1.177
## LastAuthorFemale  1.383 1      1.176
## UniqueAuthors    1.719 4      1.070
## Year             1.862 16     1.020
```


Residuals from first and last author and team size



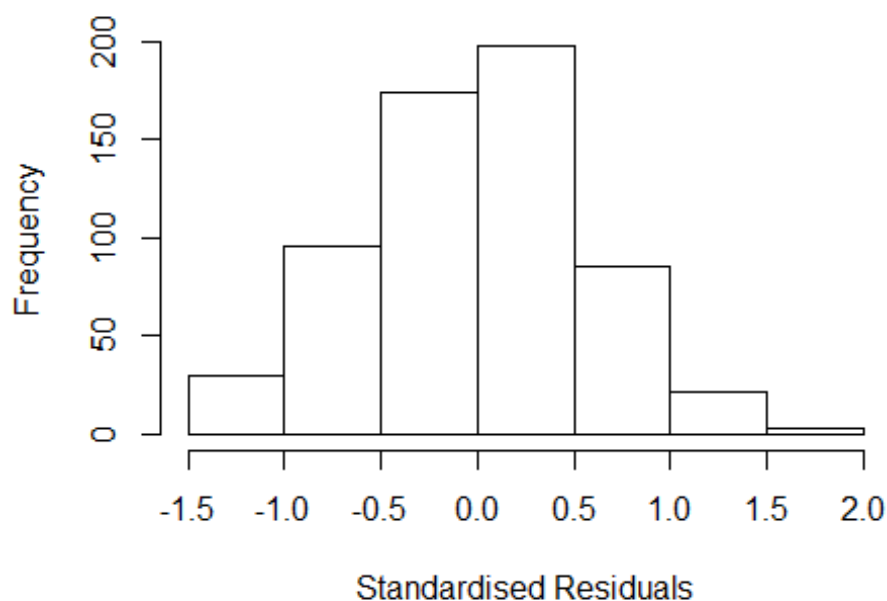
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6696 -0.3872 0.0152 0.3791 1.7674
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16365 0.12684 9.17 < 2e-16 ***
## FirstAuthorFemale1 0.00563 0.07359 0.08 0.93907
## LastAuthorFemale1 0.02648 0.09224 0.29 0.77420
## UniqueAuthors2 0.23952 0.06075 3.94 9.0e-05 ***
## UniqueAuthors3 0.30197 0.06896 4.38 1.4e-05 ***
## UniqueAuthors4 0.29518 0.08489 3.48 0.00054 ***
## UniqueAuthors5 0.38488 0.10985 3.50 0.00049 ***
## Year1997 -0.31709 0.15931 -1.99 0.04702 *
## Year1998 -0.31142 0.18316 -1.70 0.08961 .
## Year1999 -0.29729 0.19058 -1.56 0.11931
```

```

## Year2000      -0.15418    0.18150   -0.85  0.39597
## Year2001      -0.12275    0.16322   -0.75  0.45231
## Year2002      -0.28356    0.18063   -1.57  0.11699
## Year2003       0.20403    0.26697    0.76  0.44503
## Year2004      -0.28457    0.22647   -1.26  0.20941
## Year2005      -0.34219    0.16793   -2.04  0.04203 *
## Year2006      -0.11624    0.15360   -0.76  0.44949
## Year2007      -0.32446    0.15107   -2.15  0.03215 *
## Year2008      -0.25388    0.14311   -1.77  0.07658 .
## Year2009      -0.25434    0.14136   -1.80  0.07250 .
## Year2010      -0.13996    0.14036   -1.00  0.31911
## Year2011      -0.38014    0.14467   -2.63  0.00882 **
## Year2012      -0.25210    0.13983   -1.80  0.07192 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.566
## Multiple R-squared:  0.0876, Adjusted R-squared:  0.0533
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 567 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.880  0.952  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      1.64e-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 1.317 1      1.148
## LastAuthorFemale  1.347 1      1.161
## Year              1.175 16      1.005

```

Residuals from first and last author

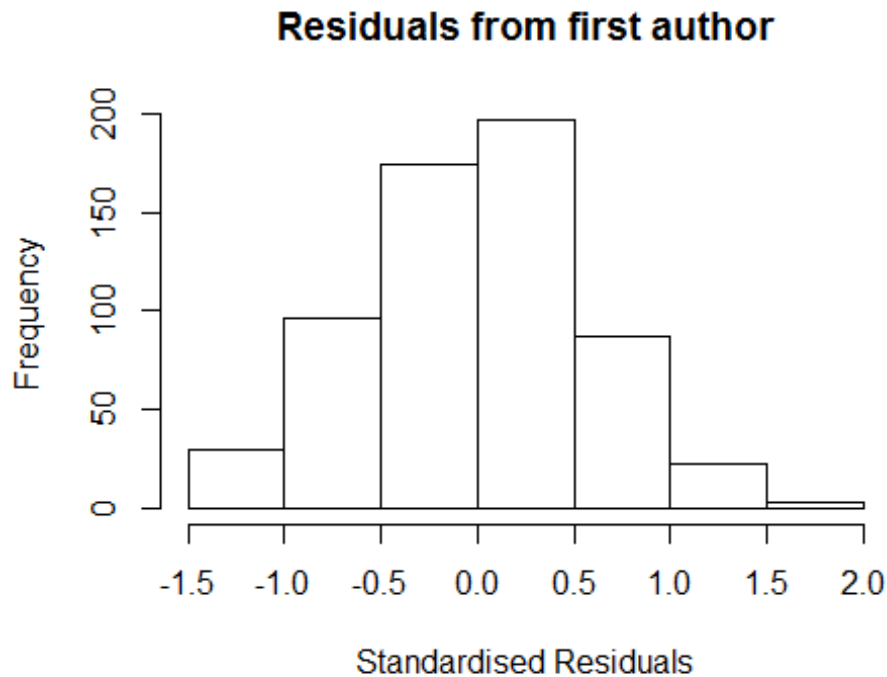


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43024 -0.40416 0.00489 0.38076 1.80454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3112 0.1211 10.83 <2e-16 ***
## FirstAuthorFemale1 0.0494 0.0742 0.67 0.506
## LastAuthorFemale1 0.0201 0.0945 0.21 0.832
## Year1997 -0.3575 0.1594 -2.24 0.025 *
## Year1998 -0.2889 0.1859 -1.55 0.121
## Year1999 -0.3106 0.1917 -1.62 0.106
## Year2000 -0.0715 0.1929 -0.37 0.711
## Year2001 -0.1059 0.1645 -0.64 0.520
## Year2002 -0.2287 0.1826 -1.25 0.211
## Year2003 0.1191 0.2542 0.47 0.640
## Year2004 -0.2972 0.2063 -1.44 0.150
## Year2005 -0.2998 0.1720 -1.74 0.082 .
```

```

## Year2006          -0.0984      0.1547   -0.64    0.525
## Year2007          -0.3280      0.1563   -2.10    0.036 *
## Year2008          -0.1968      0.1404   -1.40    0.162
## Year2009          -0.2338      0.1410   -1.66    0.098 .
## Year2010          -0.0861      0.1379   -0.62    0.533
## Year2011          -0.3294      0.1419   -2.32    0.021 *
## Year2012          -0.2011      0.1382   -1.46    0.146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0352, Adjusted R-squared:  0.00576
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 557 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.320  0.873   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      1.64e-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 1.063 1      1.031
## Year              1.063 16      1.002

```

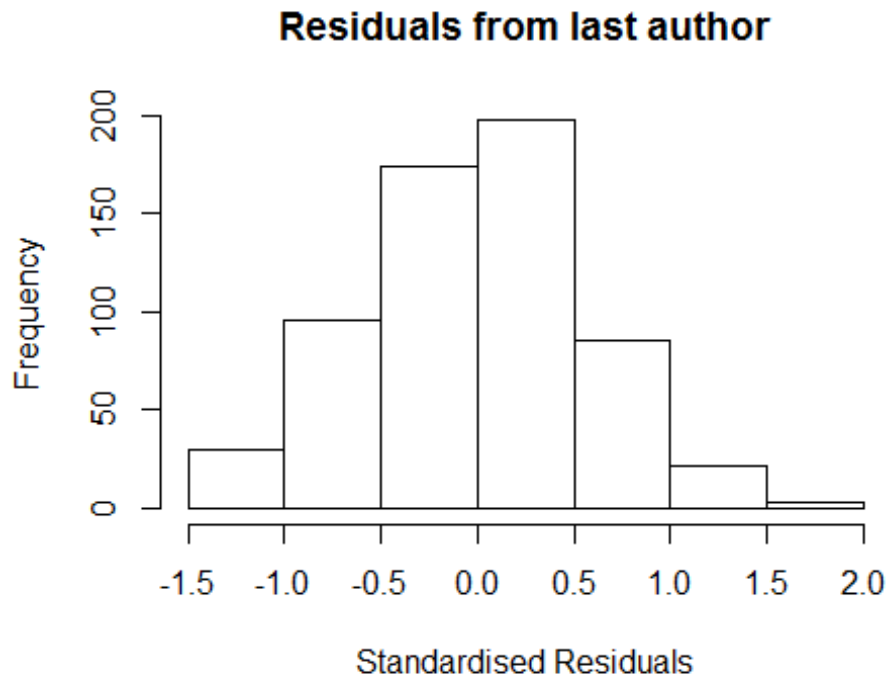


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42819 -0.40732 0.00502 0.38036 1.80124
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3128 0.1204 10.90 <2e-16 ***
## FirstAuthorFemale1 0.0557 0.0665 0.84 0.403
## Year1997 -0.3569 0.1594 -2.24 0.026 *
## Year1998 -0.2897 0.1855 -1.56 0.119
## Year1999 -0.3118 0.1908 -1.63 0.103
## Year2000 -0.0756 0.1918 -0.39 0.693
## Year2001 -0.1078 0.1639 -0.66 0.511
## Year2002 -0.2271 0.1829 -1.24 0.215
## Year2003 0.1154 0.2537 0.45 0.649
## Year2004 -0.2945 0.2058 -1.43 0.153
## Year2005 -0.3021 0.1712 -1.76 0.078 .
## Year2006 -0.0998 0.1544 -0.65 0.518
```

```

## Year2007          -0.3296      0.1553   -2.12    0.034 *
## Year2008          -0.1979      0.1396   -1.42    0.157
## Year2009          -0.2336      0.1406   -1.66    0.097 .
## Year2010          -0.0862      0.1375   -0.63    0.531
## Year2011          -0.3288      0.1416   -2.32    0.021 *
## Year2012          -0.2001      0.1379   -1.45    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0347, Adjusted R-squared:  0.00689
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 557 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.337  0.877  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      1.64e-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 1.088 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
## -1.43754 -0.40431 0.00762 0.38008 1.80162
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3153 0.1208 10.89 <2e-16 ***
## LastAuthorFemale1 0.0396 0.0849 0.47 0.641
## Year1997 -0.3605 0.1592 -2.26 0.024 *
## Year1998 -0.2856 0.1839 -1.55 0.121
## Year1999 -0.3114 0.1919 -1.62 0.105
## Year2000 -0.0663 0.1937 -0.34 0.732
## Year2001 -0.1066 0.1642 -0.65 0.516
## Year2002 -0.2299 0.1812 -1.27 0.205
## Year2003 0.1223 0.2551 0.48 0.632
## Year2004 -0.2881 0.2062 -1.40 0.163
## Year2005 -0.2996 0.1727 -1.73 0.083 .
## Year2006 -0.0999 0.1543 -0.65 0.518
```

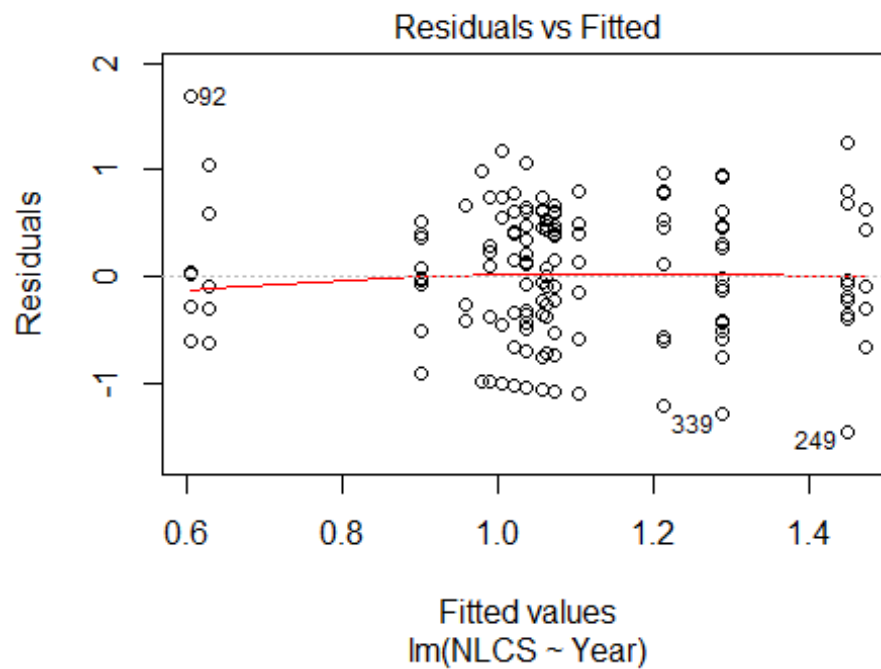
```

## Year2007          -0.3226      0.1564    -2.06      0.040 *
## Year2008          -0.1944      0.1404    -1.38      0.167
## Year2009          -0.2344      0.1413    -1.66      0.098 .
## Year2010          -0.0836      0.1375    -0.61      0.544
## Year2011          -0.3288      0.1423    -2.31      0.021 *
## Year2012          -0.2031      0.1377    -1.48      0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.00677
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 556 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.325  0.872  0.949  0.911  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      1.64e-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: 609"
## [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
##   18   33   16   25   15   16   14   10   15   20   23   26   39   29   24
## 2011 2012
##   14   29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2   14    6    7    7    6    5    3    6   10    9   12   18   17   10
## 2011 2012

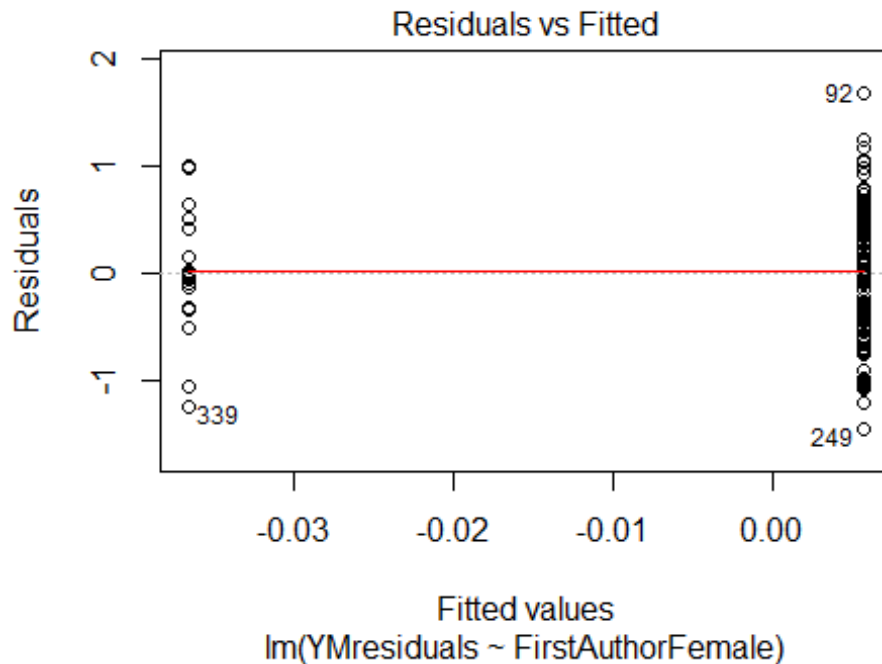
```



```
##      9    15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      1    13     6     6     5     4     4     3     6     8     8    10    13    15    10
## 2011 2012
##      8    14
## [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 = 0.24, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: 1.6013288855577"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.636e+00 1          1.624
## LastAuthorFemale  2.562e+00 1          1.601
## UniqueAuthors    2.541e+14 4          63.186
## Year              4.578e+14 16          2.872

## [1] "List of 0 outliers with 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
## -1.4507 -0.3929  0.0602  0.3799  1.9451
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```

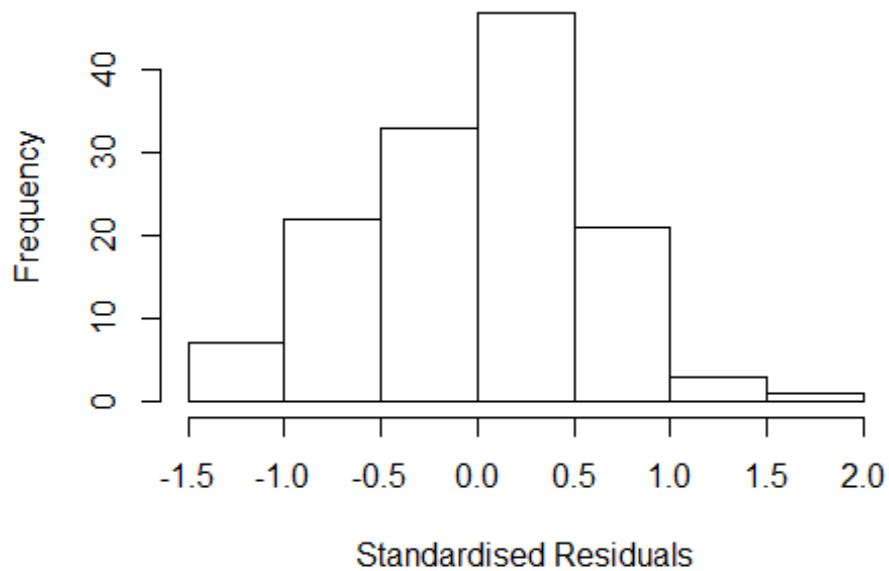
## (Intercept)          -0.291      0.165    -1.76  0.08061 .
## FirstAuthorFemale1   -0.242      0.189    -1.28  0.20241
## LastAuthorFemale1     0.197      0.197     1.00  0.31944
## UniqueAuthors2        0.279      0.148     1.89  0.06154 .
## UniqueAuthors3        0.291      0.165     1.76  0.08061 .
## UniqueAuthors4        0.352      0.188     1.87  0.06448 .
## UniqueAuthors5       -0.674      0.270    -2.49  0.01414 *
## Year1997              1.311      0.208     6.31  5.9e-09 ***
## Year1998              0.757      0.302     2.50  0.01382 *
## Year1999              0.637      0.306     2.08  0.03965 *
## Year2000              1.233      0.401     3.08  0.00263 **
## Year2001              1.082      0.872     1.24  0.21730
## Year2002              1.561      0.290     5.39  4.1e-07 ***
## Year2003              0.946      0.311     3.05  0.00288 **
## Year2004              1.129      0.245     4.61  1.1e-05 ***
## Year2005              1.741      0.294     5.93  3.6e-08 ***
## Year2006              1.390      0.238     5.84  5.2e-08 ***
## Year2007              1.571      0.280     5.61  1.5e-07 ***
## Year2008              1.463      0.222     6.60  1.5e-09 ***
## Year2009              1.175      0.195     6.04  2.1e-08 ***
## Year2010              1.226      0.189     6.47  2.7e-09 ***
## Year2011              1.170      0.294     3.98  0.00012 ***
## Year2012              0.981      0.158     6.20  9.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.629
## Multiple R-squared:  0.218, Adjusted R-squared:  0.0629
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 123 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.318  0.887  0.954   0.918  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-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01              5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)

```

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

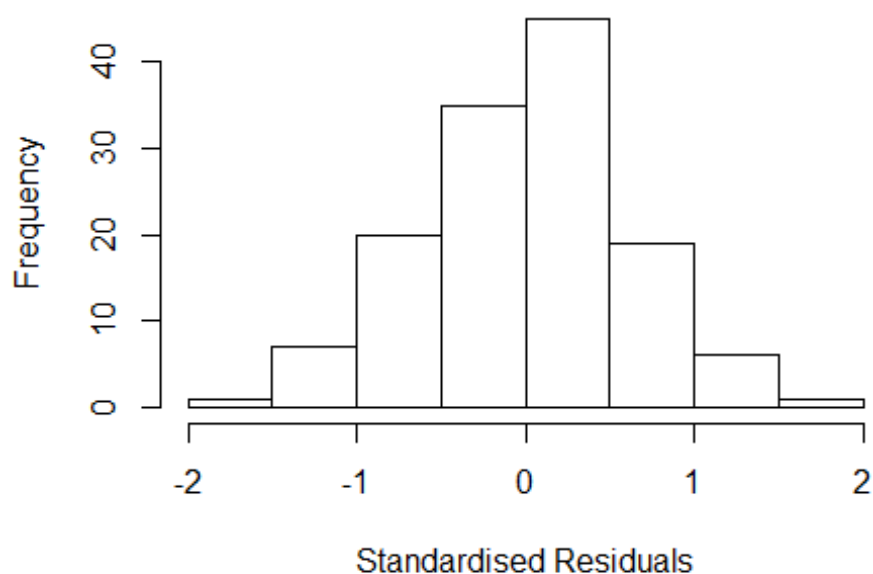
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```

Residuals from first and last author and team size



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.257  1      1.502
## LastAuthorFemale  2.152  1      1.467
## Year              3.935 16      1.044
```

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
## -1.6772 -0.4670 0.0644 0.4497 1.8137
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.77e-15 0.00e+00 Inf < 2e-16 ***
## FirstAuthorFemale1 -1.89e-01 1.88e-01 -1.00 0.3184
## LastAuthorFemale1 2.17e-01 2.19e-01 0.99 0.3219
## Year1997 1.12e+00 2.00e-01 5.60 1.5e-07 ***
## Year1998 6.31e-01 2.97e-01 2.13 0.0357 *
## Year1999 4.77e-01 2.54e-01 1.88 0.0629 .
## Year2000 1.18e+00 3.58e-01 3.29 0.0013 **
## Year2001 8.82e-01 8.33e-01 1.06 0.2918
## Year2002 1.50e+00 3.08e-01 4.88 3.5e-06 ***
## Year2003 9.40e-01 3.00e-01 3.13 0.0022 **
## Year2004 1.05e+00 2.26e-01 4.65 9.1e-06 ***
## Year2005 1.68e+00 3.12e-01 5.38 4.0e-07 ***
```

```

## Year2006          1.20e+00    2.40e-01    5.01  2.0e-06 ***
## Year2007          1.30e+00    2.69e-01    4.81  4.6e-06 ***
## Year2008          1.34e+00    1.75e-01    7.63  7.3e-12 ***
## Year2009          1.04e+00    1.59e-01    6.53  1.8e-09 ***
## Year2010          1.07e+00    1.24e-01    8.56  5.7e-14 ***
## Year2011          9.48e-01    2.25e-01    4.21  5.2e-05 ***
## Year2012          8.62e-01    1.29e-01    6.68  8.8e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.17, Adjusted R-squared:  0.0405
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~ = 1. The remaining 126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.402  0.900  0.951  0.918  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      7.46e-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 2.066 1      1.437
## Year              2.066 16      1.023
##
## [1] "List of 0 outliers 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

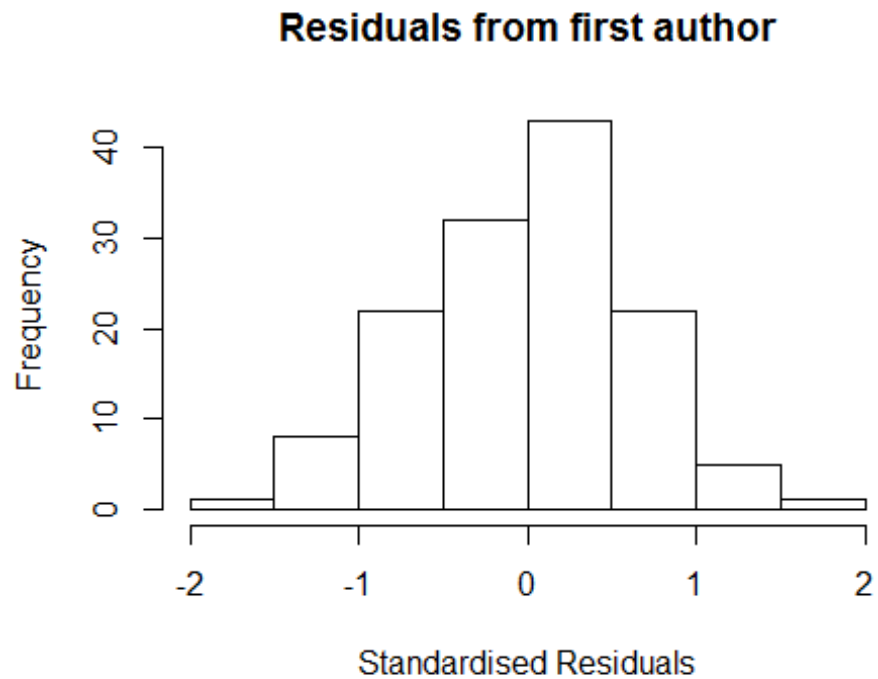
```

```

## -1.647 -0.476 0.064 0.448 1.806
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.33e-15   3.81e-08   0.00  1.0000
## FirstAuthorFemale1 -1.37e-01  1.77e-01  -0.78  0.4395
## Year1997        1.13e+00  2.01e-01   5.65  1.2e-07 ***
## Year1998        6.21e-01  2.93e-01   2.12  0.0359 *
## Year1999        4.85e-01  2.59e-01   1.87  0.0633 .
## Year2000        1.17e+00  3.64e-01   3.21  0.0017 **
## Year2001        8.86e-01  8.06e-01   1.10  0.2739
## Year2002        1.56e+00  2.69e-01   5.82  5.4e-08 ***
## Year2003        9.41e-01  2.99e-01   3.14  0.0021 **
## Year2004        1.04e+00  2.28e-01   4.55  1.3e-05 ***
## Year2005        1.65e+00  3.12e-01   5.27  6.2e-07 ***
## Year2006        1.22e+00  2.28e-01   5.35  4.4e-07 ***
## Year2007        1.33e+00  2.70e-01   4.94  2.7e-06 ***
## Year2008        1.33e+00  1.66e-01   8.01  1.0e-12 ***
## Year2009        1.04e+00  1.57e-01   6.59  1.4e-09 ***
## Year2010        1.07e+00  1.24e-01   8.57  5.1e-14 ***
## Year2011        9.75e-01  2.36e-01   4.13  6.9e-05 ***
## Year2012        9.02e-01  1.26e-01   7.16  8.1e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.654
## Multiple R-squared:  0.161, Adjusted R-squared:  0.0385
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 125 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.425 0.901 0.951 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      7.46e-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"

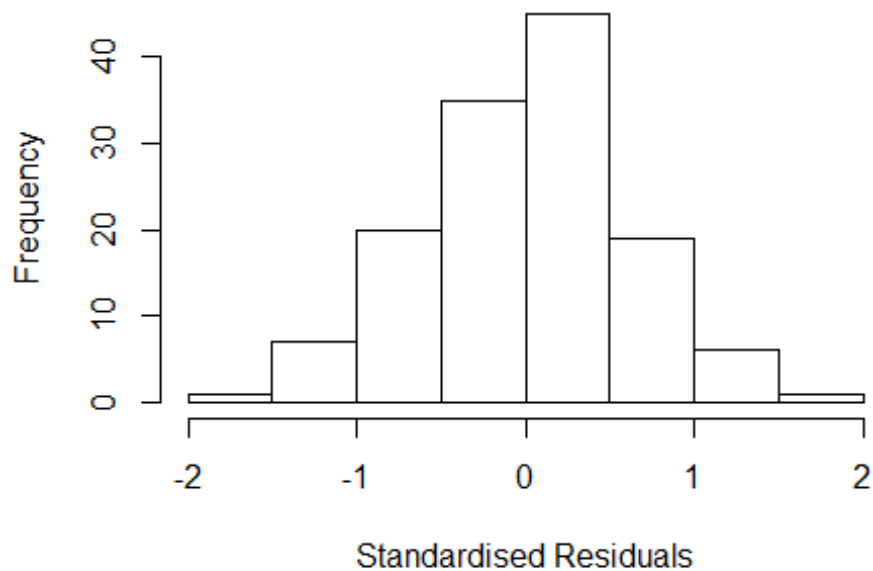
```

```
## 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 1.841 1      1.357  
## Year             1.841 16      1.019
```


Residuals from last author



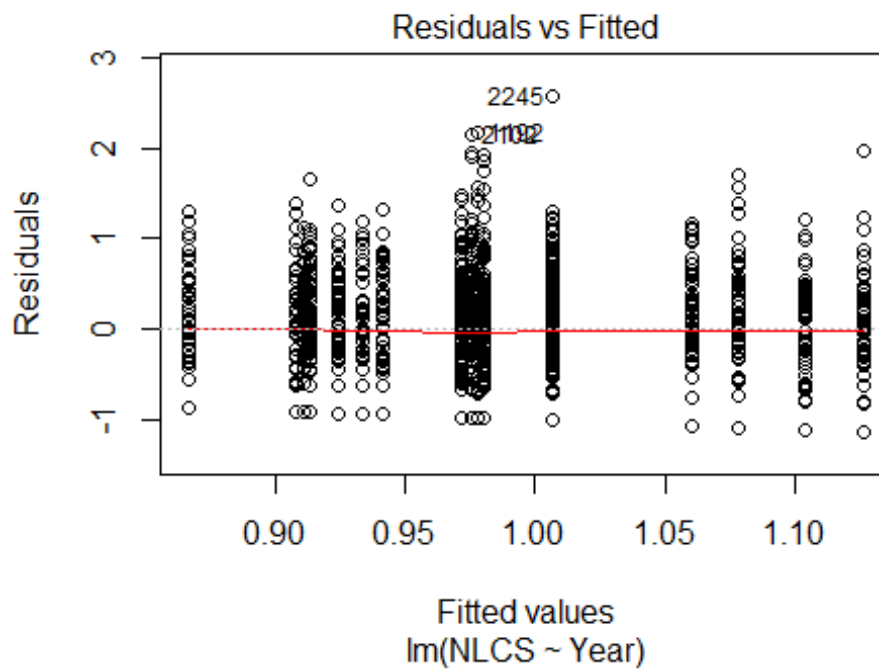
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5756 -0.4629 0.0131 0.4768 1.8057
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.61e-16 0.00e+00 Inf < 2e-16 ***
## LastAuthorFemale1 1.46e-01 2.25e-01 0.65 0.5175
## Year1997 1.11e+00 1.96e-01 5.65 1.2e-07 ***
## Year1998 5.92e-01 2.87e-01 2.06 0.0413 *
## Year1999 4.85e-01 2.58e-01 1.88 0.0624 .
## Year2000 1.16e+00 3.93e-01 2.94 0.0039 **
## Year2001 8.86e-01 8.05e-01 1.10 0.2733
## Year2002 1.52e+00 2.98e-01 5.10 1.3e-06 ***
## Year2003 9.41e-01 2.99e-01 3.15 0.0021 **
## Year2004 1.01e+00 2.35e-01 4.32 3.3e-05 ***
## Year2005 1.58e+00 2.90e-01 5.43 3.2e-07 ***
## Year2006 1.16e+00 2.34e-01 4.97 2.3e-06 ***
```

```

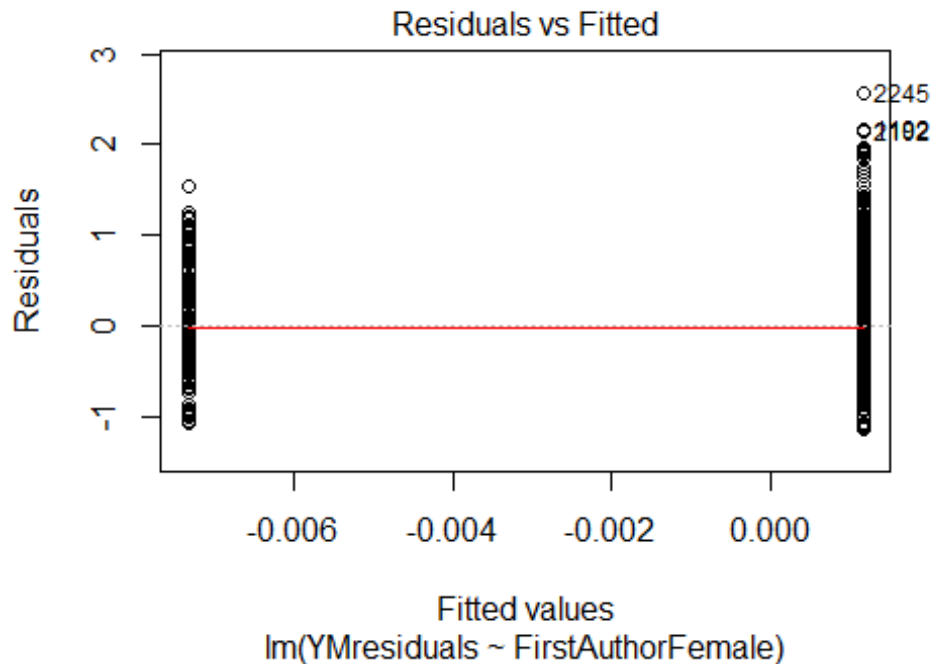
## Year2007          1.25e+00    2.67e-01    4.70  7.3e-06 ***
## Year2008          1.29e+00    1.71e-01    7.56  1.0e-11 ***
## Year2009          1.03e+00    1.54e-01    6.69  8.3e-10 ***
## Year2010          1.07e+00    1.24e-01    8.57  5.1e-14 ***
## Year2011          9.57e-01    2.28e-01    4.20  5.4e-05 ***
## Year2012          8.48e-01    1.28e-01    6.64  1.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.654
## Multiple R-squared:  0.159, Adjusted R-squared:  0.0359
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.426  0.908  0.951  0.921  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      7.46e-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: 134"
## [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
## 132 106 103 106 95 113 100 92 124 135 104 146 192 201 204
## 2011 2012
## 200 216
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 63 60 50 45 61 56 55 74 82 64 88 118 130 129
## 2011 2012

```

```
## 134 150
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 47 50 39 39 52 46 48 64 74 54 75 111 110 114
## 2011 2012
## 114 135
## [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
```

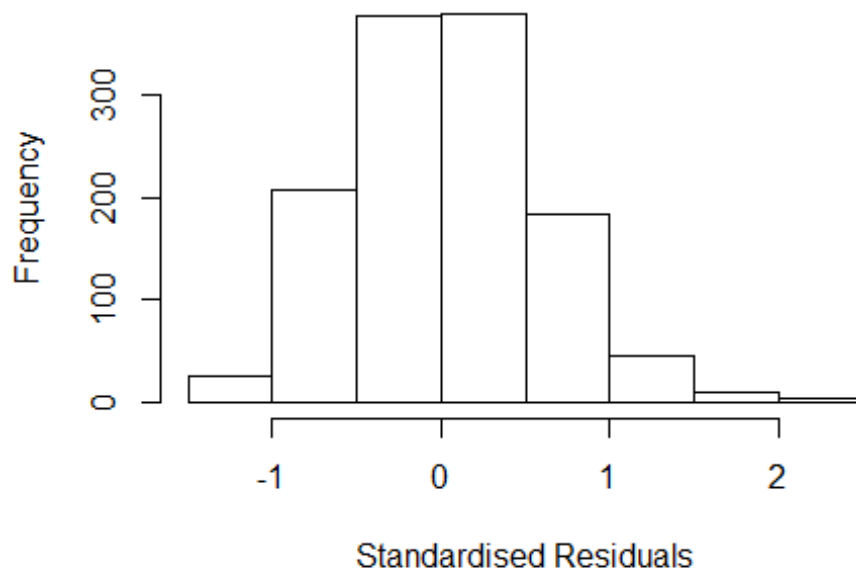


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.092, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 2.79088047621233"
## [1] "Male first author team size 2018 geometric mean: 1.90624421218627"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 920, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.50303918507579"
## [1] "Male last author team size 2018 geometric mean: 1.96581677446167"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680, 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 1.197 1      1.094
## LastAuthorFemale  1.177 1      1.085
## UniqueAuthors    1.318 4      1.035
## Year              1.392 16     1.010
```

Residuals from first and last author and team size



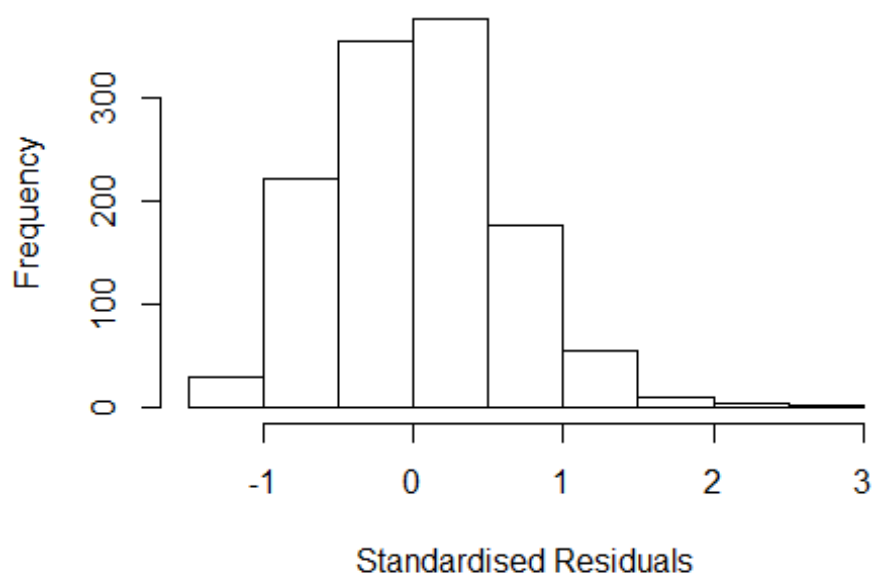
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35203 -0.39283 0.00975 0.39855 2.39885
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9690 0.0959 10.10 < 2e-16 ***
## FirstAuthorFemale1 -0.0768 0.0554 -1.39 0.16600
## LastAuthorFemale1 -0.0278 0.0531 -0.52 0.59975
## UniqueAuthors2 0.2068 0.0402 5.15 3.1e-07 ***
## UniqueAuthors3 0.3830 0.0514 7.46 1.7e-13 ***
## UniqueAuthors4 0.4585 0.0751 6.10 1.4e-09 ***
## UniqueAuthors5 0.3091 0.0835 3.70 0.00022 ***
## Year1997 -0.0146 0.1281 -0.11 0.90946
## Year1998 -0.1385 0.1291 -1.07 0.28339
## Year1999 -0.2485 0.1395 -1.78 0.07499 .
```

```

## Year2000          -0.2007      0.1377    -1.46    0.14518
## Year2001          -0.0872      0.1290    -0.68    0.49934
## Year2002          -0.0438      0.1203    -0.36    0.71548
## Year2003          -0.1784      0.1236    -1.44    0.14893
## Year2004          -0.2179      0.1300    -1.68    0.09402 .
## Year2005          -0.2444      0.1191    -2.05    0.04034 *
## Year2006          -0.1818      0.1233    -1.47    0.14069
## Year2007          -0.2112      0.1145    -1.85    0.06528 .
## Year2008          -0.2236      0.1120    -2.00    0.04610 *
## Year2009          -0.1674      0.1110    -1.51    0.13177
## Year2010          -0.1283      0.1103    -1.16    0.24482
## Year2011          -0.2132      0.1107    -1.93    0.05434 .
## Year2012          -0.1968      0.1090    -1.81    0.07130 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.0756, Adjusted R-squared:  0.0588
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0454 0.8630 0.9480 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          8.14e-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 1.160 1          1.077
## LastAuthorFemale 1.120 1          1.059
## Year          1.128 16          1.004

```

Residuals from first and last author

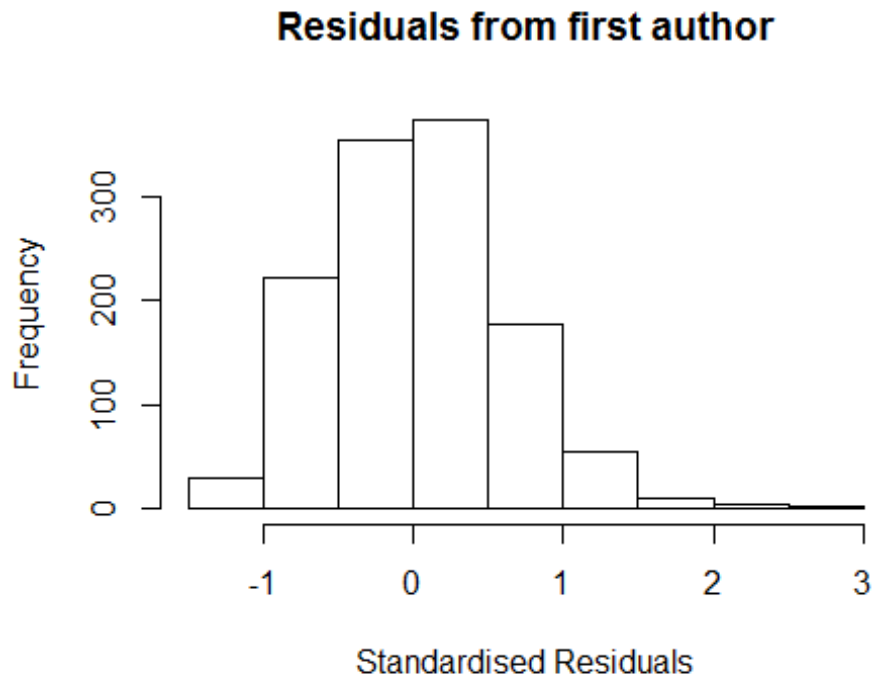


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2245 75249087100 3.574 2009      1303      6      2.596
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0923 -0.3939  0.0113  0.4037  2.5962
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07e+00   9.31e-02  11.50  <2e-16 ***
## FirstAuthorFemale1 -2.82e-02   5.67e-02  -0.50   0.619
## LastAuthorFemale1  2.18e-02   5.35e-02   0.41   0.684
## Year1997        -3.73e-05   1.28e-01   0.00   1.000
## Year1998        -1.50e-01   1.31e-01  -1.15   0.251
## Year1999        -2.85e-01   1.46e-01  -1.96   0.051 .
## Year2000        -1.28e-01   1.41e-01  -0.91   0.365
## Year2001        -1.76e-02   1.30e-01  -0.14   0.892
## Year2002        -2.78e-03   1.21e-01  -0.02   0.982
## Year2003        -1.65e-01   1.22e-01  -1.36   0.174
## Year2004        -2.05e-01   1.28e-01  -1.60   0.110
## Year2005        -1.70e-01   1.18e-01  -1.44   0.150
```

```

## Year2006      -1.47e-01  1.25e-01  -1.18  0.237
## Year2007      -1.82e-01  1.13e-01  -1.61  0.108
## Year2008      -1.68e-01  1.11e-01  -1.51  0.131
## Year2009      -9.27e-02  1.08e-01  -0.86  0.392
## Year2010      -6.41e-02  1.09e-01  -0.59  0.556
## Year2011      -1.39e-01  1.09e-01  -1.27  0.204
## Year2012      -9.36e-02  1.06e-01  -0.88  0.378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0131, Adjusted R-squared:  -0.00153
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1125 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0211 0.8640 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      8.14e-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 1.083 1      1.040
## Year      1.083 16      1.002

```

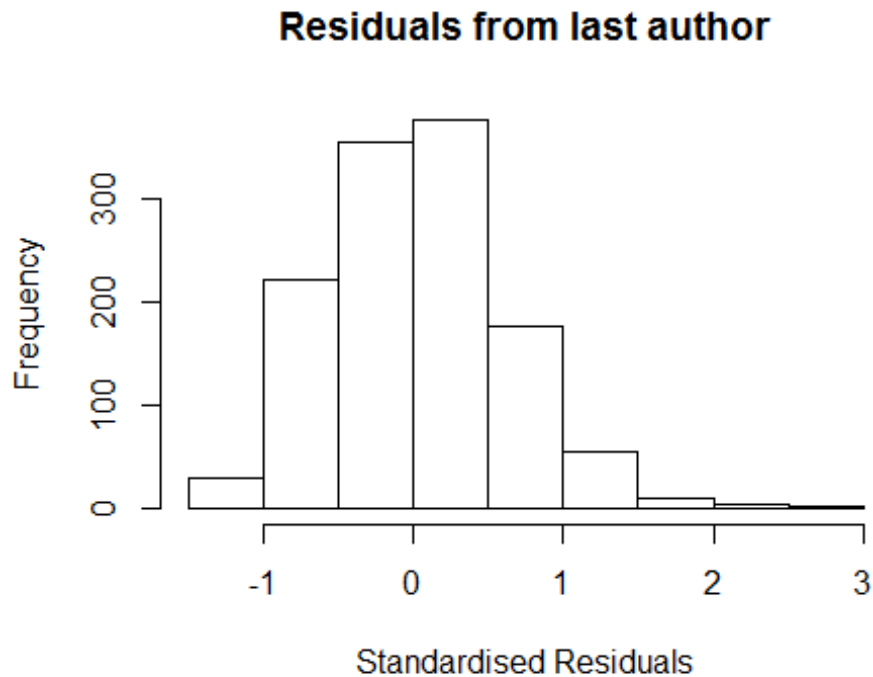



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2245 75249087100 3.574 2009      1303      6      2.596
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0717 -0.3961  0.0125  0.4029  2.5937
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.070918   0.093052   11.51  <2e-16 ***
## FirstAuthorFemale1 -0.021680   0.054671   -0.40    0.692
## Year1997         0.000805   0.128595    0.01    0.995
## Year1998        -0.148495   0.130568   -1.14    0.256
## Year1999        -0.283391   0.145575   -1.95    0.052 .
## Year2000        -0.127257   0.140725   -0.90    0.366
## Year2001        -0.017337   0.129970   -0.13    0.894
## Year2002        -0.001378   0.121243   -0.01    0.991
## Year2003        -0.162816   0.121721   -1.34    0.181
## Year2004        -0.203117   0.127946   -1.59    0.113
## Year2005        -0.170338   0.118140   -1.44    0.150
## Year2006        -0.145162   0.124414   -1.17    0.244
```

```

## Year2007          -0.180838    0.113033    -1.60     0.110
## Year2008          -0.165970    0.110817    -1.50     0.134
## Year2009          -0.090569    0.107891    -0.84     0.401
## Year2010          -0.062387    0.108529    -0.57     0.566
## Year2011          -0.136987    0.108916    -1.26     0.209
## Year2012          -0.091771    0.105913    -0.87     0.386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.013, Adjusted R-squared:  -0.000825
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1125 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0211 0.8650 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      8.14e-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 1.049 1          1.024
## Year            1.049 16          1.002

```



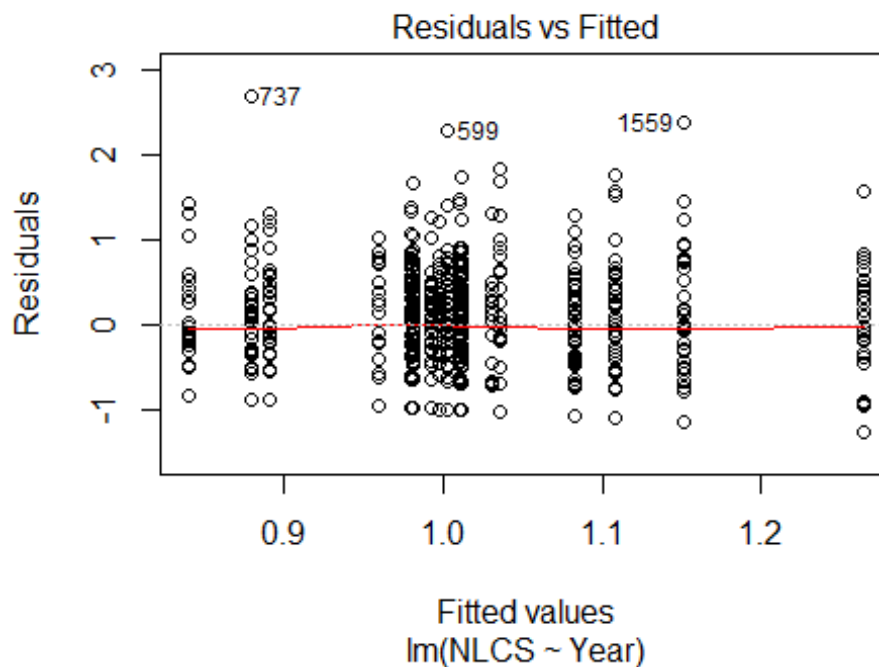
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2245 75249087100 3.574 2009      1303      6      2.596
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0817 -0.4051  0.0134  0.4034  2.5971
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.068687   0.092874   11.51  <2e-16 ***
## LastAuthorFemale1 0.013006   0.052130    0.25    0.80
## Year1997      -0.000854   0.128333   -0.01    0.99
## Year1998      -0.150767   0.130507   -1.16    0.25
## Year1999      -0.286072   0.145616   -1.96    0.05 *
## Year2000      -0.129434   0.140760   -0.92    0.36
## Year2001      -0.017379   0.129908   -0.13    0.89
## Year2002      -0.005504   0.120648   -0.05    0.96
## Year2003      -0.166710   0.121675   -1.37    0.17
## Year2004      -0.205874   0.127859   -1.61    0.11
## Year2005      -0.170682   0.117898   -1.45    0.15
## Year2006      -0.145925   0.124414   -1.17    0.24
```

```

## Year2007          -0.184621    0.112910    -1.64     0.10
## Year2008          -0.167180    0.110950    -1.51     0.13
## Year2009          -0.091760    0.108224    -0.85     0.40
## Year2010          -0.064041    0.108734    -0.59     0.56
## Year2011          -0.141473    0.108624    -1.30     0.19
## Year2012          -0.094978    0.105905    -0.90     0.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.6
## Multiple R-squared:  0.013, Adjusted R-squared:  -0.000904
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0215 0.8640 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      8.14e-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: 1229"
## [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
##   67   73   54   62   65   64   64   47   58   41   69   76   86  103   61
## 2011 2012
##   72   70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   37   25   28   32   30   40   39   19   39   33   43   51   62   59   38
## 2011 2012

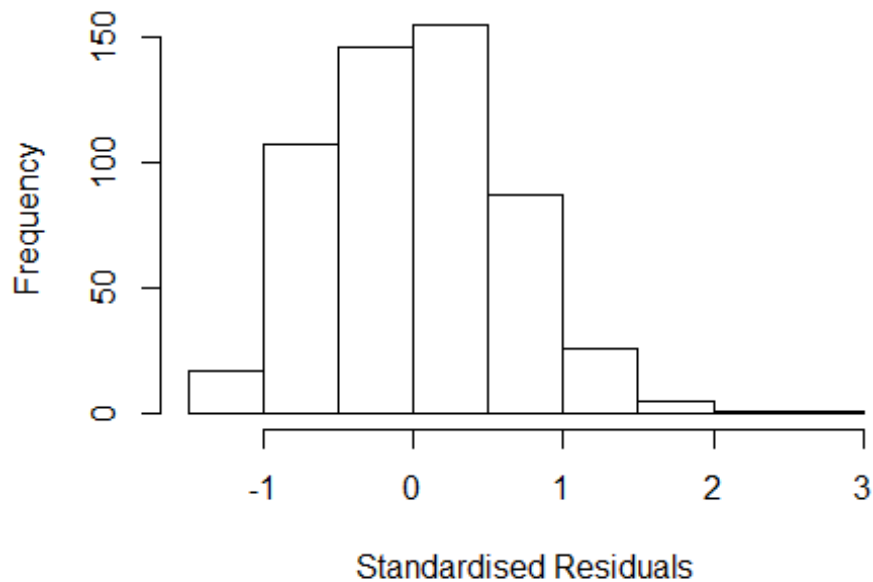
```

```
## 47 40
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 23 24 24 23 32 26 17 35 31 34 44 51 51 29
## 2011 2012
## 40 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 = 21, df = 16, p-value = 0.2
```



```
##
## 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"
## [1] "Male first author team size 2018 geometric mean: 1.9386676547219"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```


Residuals from first and last author and team size



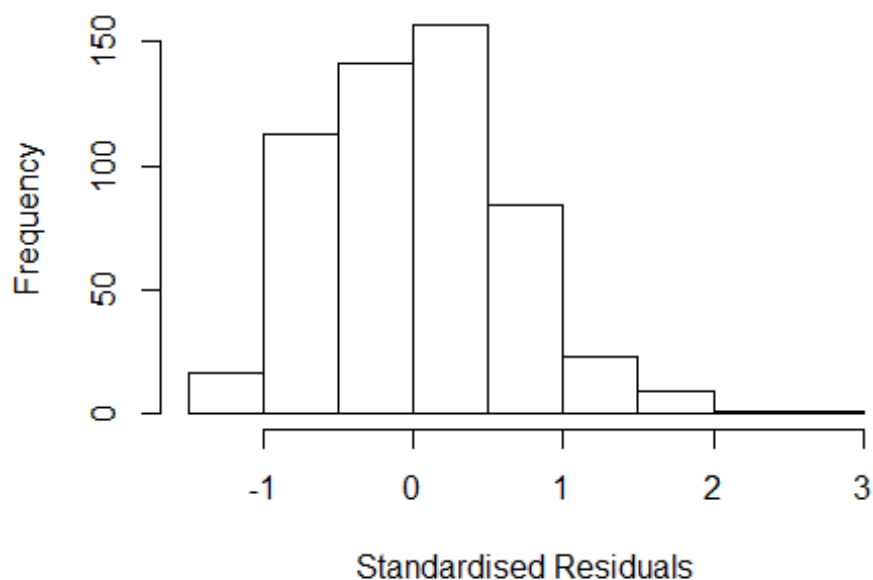
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 737 3142768423 3.571 2004      1703      3      2.569
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.28491 -0.45549  0.00482  0.43854  2.56876
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.860063   0.155399   5.53 4.9e-08 ***
## FirstAuthorFemale1 0.139749   0.079469   1.76 0.07924 .
## LastAuthorFemale1 -0.083659   0.072322  -1.16 0.24790
## UniqueAuthors2    0.050899   0.068616   0.74 0.45855
## UniqueAuthors3    0.269111   0.075770   3.55 0.00042 ***
## UniqueAuthors4    0.289055   0.177341   1.63 0.10372
## UniqueAuthors5    0.637958   0.260743   2.45 0.01475 *
## Year1997          0.037389   0.218137   0.17 0.86398
## Year1998          0.234198   0.278553   0.84 0.40086
## Year1999         -0.155324   0.188031  -0.83 0.40915
```

```

## Year2000      0.232149    0.219274    1.06    0.29022
## Year2001     -0.061397    0.206308   -0.30    0.76613
## Year2002     -0.069746    0.208133   -0.34    0.73768
## Year2003      0.074727    0.208628    0.36    0.72035
## Year2004     -0.126935    0.181239   -0.70    0.48401
## Year2005      0.040919    0.180661    0.23    0.82091
## Year2006      0.045396    0.191534    0.24    0.81274
## Year2007     -0.026808    0.177444   -0.15    0.87997
## Year2008     -0.098947    0.180020   -0.55    0.58280
## Year2009     -0.000909    0.171004   -0.01    0.99576
## Year2010      0.152831    0.187138    0.82    0.41449
## Year2011      0.127299    0.180417    0.71    0.48076
## Year2012      0.121011    0.200724    0.60    0.54686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.671
## Multiple R-squared:  0.0622, Adjusted R-squared:  0.0227
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 506 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.878  0.952  0.919  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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.151 1      1.073
## LastAuthorFemale  1.200 1      1.096
## Year              1.235 16      1.007

```


Residuals from first and last author



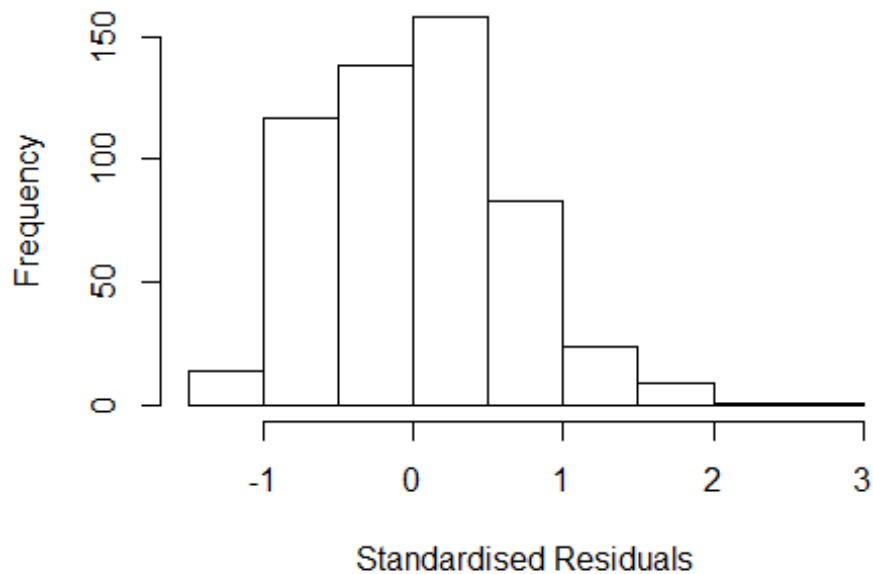
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 737 3142768423 3.571 2004      1703      3      2.757
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26035 -0.46164  0.00764  0.42936  2.75736
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8994    0.1481    6.07 2.4e-09 ***
## FirstAuthorFemale1  0.1313    0.0779    1.69  0.093 .
## LastAuthorFemale1 -0.0621    0.0695   -0.89  0.372
## Year1997         0.0457    0.2125    0.22  0.830
## Year1998         0.2297    0.2683    0.86  0.392
## Year1999        -0.1527    0.1843   -0.83  0.408
## Year2000         0.2770    0.2184    1.27  0.205
## Year2001        -0.0261    0.1966   -0.13  0.895
## Year2002        -0.0378    0.2048   -0.18  0.854
## Year2003         0.0859    0.2014    0.43  0.670
## Year2004        -0.0857    0.1791   -0.48  0.632
## Year2005         0.0683    0.1791    0.38  0.703
```

```

## Year2006          0.1016      0.1841      0.55      0.581
## Year2007          0.0227      0.1740      0.13      0.896
## Year2008         -0.0048      0.1729     -0.03      0.978
## Year2009          0.0784      0.1658      0.47      0.637
## Year2010          0.1798      0.1811      0.99      0.321
## Year2011          0.1676      0.1746      0.96      0.338
## Year2012          0.1917      0.2070      0.93      0.355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.677
## Multiple R-squared:  0.0284, Adjusted R-squared:  -0.00481
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 498 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0598 0.8770 0.9520 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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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      1.041
## Year              1.085 16      1.003

```

Residuals from first author



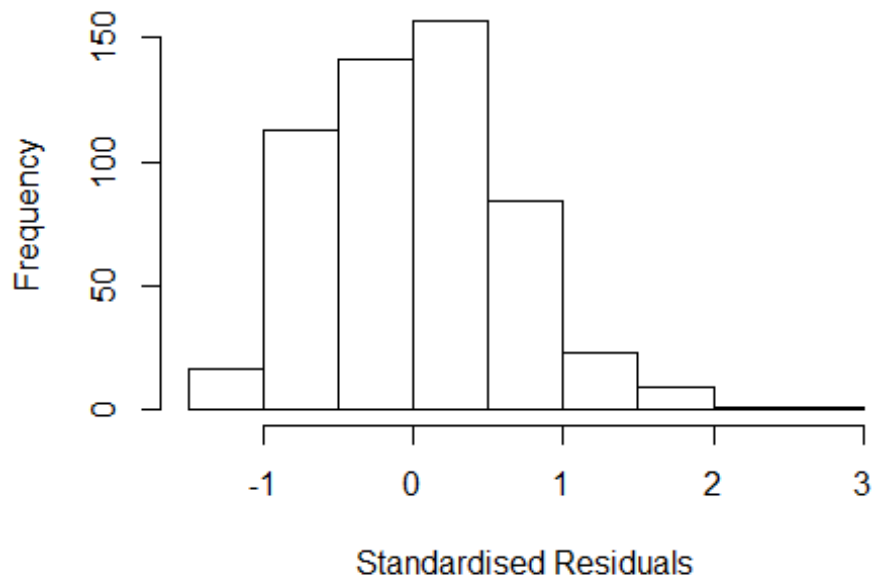
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 737 3142768423 3.571 2004      1703      3      2.757
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23773 -0.45011  0.00641  0.43319  2.77095
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.89370    0.14766   6.05 2.7e-09 ***
## FirstAuthorFemale1 0.10714    0.07589   1.41  0.16
## Year1997        0.04387    0.21156   0.21  0.84
## Year1998        0.23689    0.26767   0.89  0.38
## Year1999       -0.15688    0.18274  -0.86  0.39
## Year2000        0.27682    0.21727   1.27  0.20
## Year2001       -0.02481    0.19631  -0.13  0.90
## Year2002       -0.03925    0.20436  -0.19  0.85
## Year2003        0.08538    0.20125   0.42  0.67
## Year2004       -0.09364    0.17846  -0.52  0.60
## Year2005        0.06941    0.17885   0.39  0.70
## Year2006        0.10308    0.18389   0.56  0.58
```

```

## Year2007          0.02162      0.17372      0.12      0.90
## Year2008         -0.00767      0.17228     -0.04      0.96
## Year2009          0.08299      0.16545      0.50      0.62
## Year2010          0.17105      0.17946      0.95      0.34
## Year2011          0.16313      0.17377      0.94      0.35
## Year2012          0.18742      0.20638      0.91      0.36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.677
## Multiple R-squared:  0.0274, Adjusted R-squared:  -0.00397
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 506 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0562 0.8770 0.9520 0.9180 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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.131 1      1.064
## Year              1.131 16      1.004

```

Residuals from last author



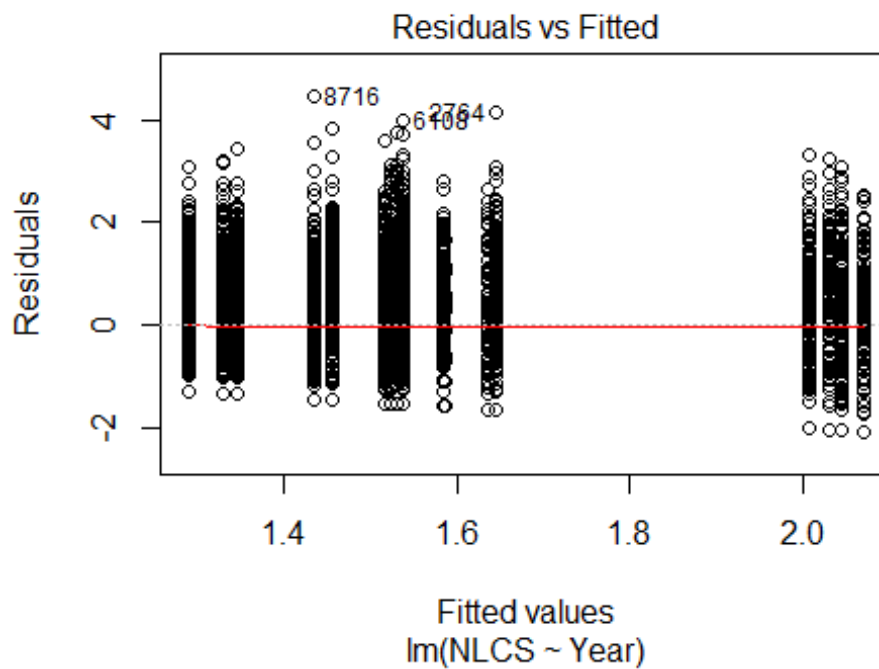
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 737 3142768423 3.571 2004      1703      3      2.757
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1724 -0.4512  0.0129  0.4328  2.7482
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90699    0.15150   5.99   4e-09 ***
## LastAuthorFemale1 -0.01960    0.06938  -0.28    0.78
## Year1997         0.04487    0.21663   0.21    0.84
## Year1998         0.23116    0.26566   0.87    0.38
## Year1999        -0.14567    0.18722  -0.78    0.44
## Year2000         0.26538    0.22021   1.21    0.23
## Year2001        -0.02019    0.20080  -0.10    0.92
## Year2002        -0.04564    0.20645  -0.22    0.83
## Year2003         0.08009    0.20241   0.40    0.69
## Year2004        -0.08419    0.18192  -0.46    0.64
## Year2005         0.07032    0.18328   0.38    0.70
## Year2006         0.10317    0.18793   0.55    0.58
```

```

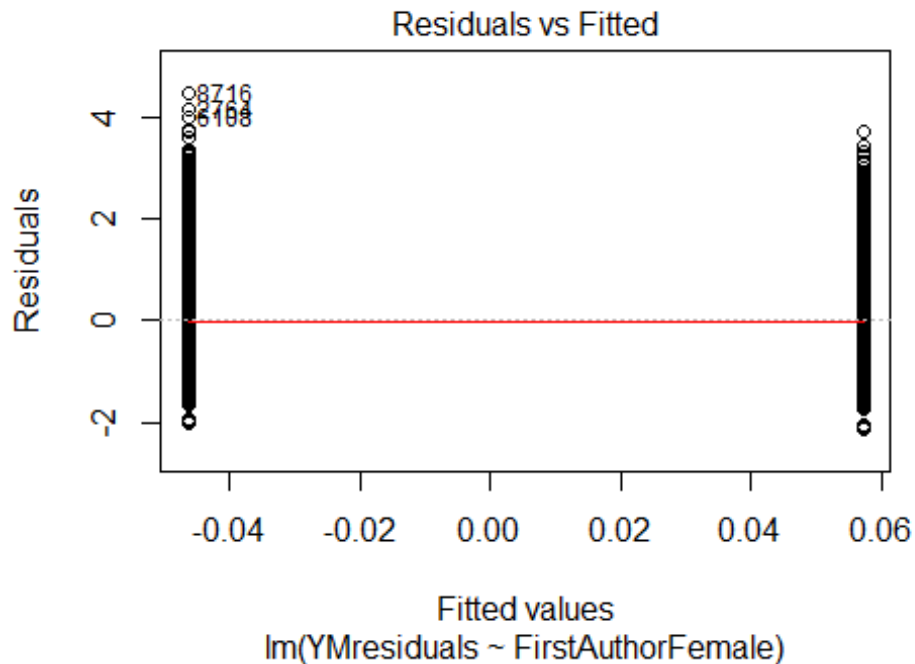
## Year2007          0.03925      0.17713      0.22      0.82
## Year2008          0.00333      0.17645      0.02      0.98
## Year2009          0.08721      0.16949      0.51      0.61
## Year2010          0.17508      0.18447      0.95      0.34
## Year2011          0.15951      0.17769      0.90      0.37
## Year2012          0.20260      0.20843      0.97      0.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.677
## Multiple R-squared:  0.0247, Adjusted R-squared:  -0.00673
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 502 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0623 0.8770 0.9530 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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 545"
## [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
## 333 378 397 405 50 546 508 469 544 569 726 1014 971 979 1136
## 2011 2012
## 1069 1255
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 181 220 208 213 17 109 371 359 419 470 558 827 787 769 886
## 2011 2012

```

```
## 839 984
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 163 197 191 193 16 100 347 332 388 421 511 750 712 673 796
## 2011 2012
## 764 903
## [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
```

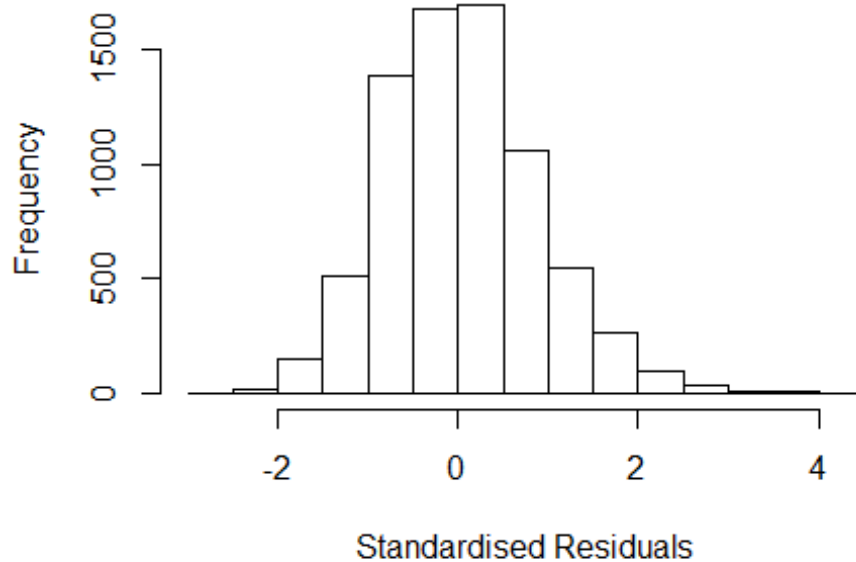


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.7, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.7614059327417"
## [1] "Male first author team size 2018 geometric mean: 3.23047588357711"
##
## 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: 3.50046204067266"
## [1] "Male last author team size 2018 geometric mean: 3.54258965827311"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 29000, 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 1.197 1      1.094
## LastAuthorFemale  1.181 1      1.087
## UniqueAuthors    1.082 4      1.010
## Year              1.105 16     1.003
```


Residuals from first and last author and team size



```
## [1] "List of 50 outliers with residuals above 2.5"
##          ScopusId  NLCS Year OneField Fields residuals
## 41      0030993455 0.000 1996      2700      1    -2.653
## 233     0029880591 3.869 1996      2700      1     2.543
## 243     13344270907 0.000 1996      2700      1    -2.627
## 262     0030032146 5.242 1996      2700      1     3.006
## 471     0030764023 4.935 1997      2700      1     2.574
## 493     0030835927 4.545 1997      2700      1     2.593
## 579     0030997937 4.151 1997      2700      1     2.830
## 784     84920298850 0.000 1997      2700      1    -2.649
## 1590    0033585476 5.332 1999      2700      1     3.074
## 1973    0034913890 3.625 2001      2700      1     2.712
## 2152    0035901121 3.724 2001      2700      1     2.811
## 2451    0037020941 3.405 2002      2700      1     2.521
## 2738    0037035126 4.738 2002      2700      1     2.766
## 2764    0037148929 5.770 2002      2700      1     3.563
## 2940    0344666767 5.279 2003      2700      1     3.197
## 3172    0037685217 4.392 2003      2700      1     2.545
## 3765    2942709630 4.228 2004      2700      1     2.772
## 4001    0842265438 4.398 2004      2700      1     2.609
## 4004    0742304414 3.406 2004      2700      1     2.531
## 4447    20444443626 4.732 2005      2700      1     4.009
## 4449    20444442419 5.269 2005      2700      1     3.219
## 4545    17244374336 3.227 2005      2700      1     2.504
## 5166    33748472277 3.462 2006      2700      1     2.665
## 5443    33645077861 4.460 2006      2700      1     3.663
## 5524    30944466084 4.625 2006      2700      1     2.527
```

```

## 5879 34948835806 4.859 2007 2700 1 2.710
## 5887 38149021049 3.639 2007 2700 1 2.818
## 6016 34548294922 4.772 2007 2700 1 2.627
## 6022 34548047185 5.239 2007 2700 1 3.117
## 6034 34548152773 3.324 2007 2700 1 2.503
## 6108 34347323902 5.512 2007 2700 1 3.629
## 6178 34447278870 3.548 2007 2700 1 2.727
## 6287 34248140087 4.289 2007 2700 1 2.580
## 6627 33846543092 4.549 2007 2700 1 2.661
## 6764 55949095116 3.528 2008 2700 1 2.733
## 7529 41649095977 4.014 2008 2700 1 3.269
## 7736 38349136266 5.093 2008 2700 1 3.024
## 7862 72449172862 4.437 2009 2700 1 2.523
## 8283 68249144371 4.965 2009 2700 1 3.001
## 8658 65249183059 3.696 2009 2700 1 2.502
## 8716 63249128249 5.883 2009 2700 1 3.969
## 9175 78650205262 3.972 2010 2700 1 2.577
## 9949 77952306239 4.487 2010 2700 1 2.678
## 10035 77951046282 4.541 2010 2700 1 2.994
## 10670 80555126201 3.058 2011 2700 1 2.536
## 10699 80054991418 3.610 2011 2700 1 2.507
## 10752 80054955347 4.759 2011 2700 1 2.910
## 11795 78651385897 3.245 2011 2700 1 2.723
## 12791 84865093812 2.997 2012 2700 1 2.565
## 12893 84860702705 4.347 2012 2700 1 2.591
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.65347 -0.56616 -0.00729 0.54560 4.00943
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32574 0.09942 13.34 < 2e-16 ***
## FirstAuthorFemale1 0.02680 0.02167 1.24 0.2161
## LastAuthorFemale1 0.02260 0.02207 1.02 0.3057
## UniqueAuthors2 0.58078 0.03198 18.16 < 2e-16 ***
## UniqueAuthors3 0.88743 0.03157 28.11 < 2e-16 ***
## UniqueAuthors4 1.03942 0.03265 31.83 < 2e-16 ***
## UniqueAuthors5 1.30093 0.02817 46.18 < 2e-16 ***
## Year1997 -0.00432 0.12983 -0.03 0.9734
## Year1998 -0.00242 0.12069 -0.02 0.9840
## Year1999 -0.10765 0.12334 -0.87 0.3828
## Year2000 -0.11666 0.27575 -0.42 0.6723

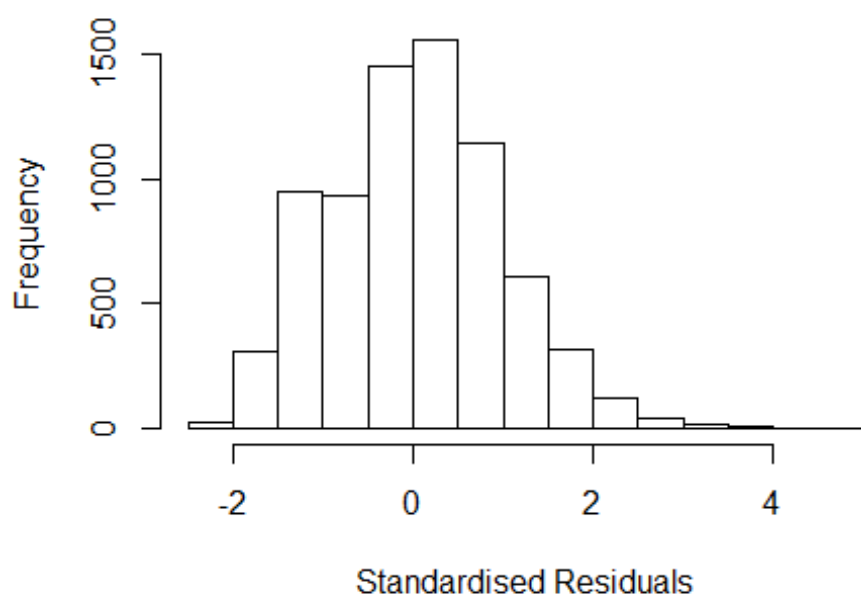
```

```

## Year2001      -0.41319      0.14686      -2.81      0.0049 **
## Year2002      -0.44219      0.11162      -3.96      7.5e-05 ***
## Year2003      -0.54464      0.11450      -4.76      2.0e-06 ***
## Year2004      -0.45076      0.10994      -4.10      4.2e-05 ***
## Year2005      -0.60317      0.11013      -5.48      4.5e-08 ***
## Year2006      -0.52916      0.10722      -4.94      8.2e-07 ***
## Year2007      -0.50446      0.10435      -4.83      1.4e-06 ***
## Year2008      -0.58028      0.10332      -5.62      2.0e-08 ***
## Year2009      -0.71249      0.10338      -6.89      5.9e-12 ***
## Year2010      -0.84457      0.10151      -8.32      < 2e-16 ***
## Year2011      -0.80399      0.10249      -7.84      4.9e-15 ***
## Year2012      -0.89359      0.10073      -8.87      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.794
## Multiple R-squared:  0.288, Adjusted R-squared:  0.286
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(2147,4830) are outliers with |weight| = 0 ( < 1.3e-05);
## 644 weights are ~ = 1. The remaining 6811 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0009 0.8670 0.9470 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          1.34e-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 1.105 1          1.051
## LastAuthorFemale 1.108 1          1.052
## Year 1.032 16          1.001

```

Residuals from first and last author



```
## [1] "List of 66 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2      0029957675 5.002 1996      2700      1      2.879
## 103     0029842750 4.652 1996      2700      1      2.675
## 262     0030032146 5.242 1996      2700      1      3.346
## 471     0030764023 4.935 1997      2700      1      2.960
## 493     0030835927 4.545 1997      2700      1      2.503
## 536     0030859363 4.598 1997      2700      1      2.623
## 636     0031041139 5.114 1997      2700      1      3.219
## 821     0032512287 4.574 1998      2700      1      2.531
## 1590    0033585476 5.332 1999      2700      1      3.397
## 1597    0033573788 4.886 1999      2700      1      2.804
## 2135    0035292711 4.056 2001      2700      1      2.604
## 2332    0035661067 4.269 2001      2700      1      2.737
## 2705    0037070806 4.592 2002      2700      1      3.083
## 2738    0037035126 4.738 2002      2700      1      3.082
## 2764    0037148929 5.770 2002      2700      1      4.261
## 2773    0037116811 4.433 2002      2700      1      2.844
## 2940    0344666767 5.279 2003      2700      1      3.894
## 2991    0345374593 4.258 2003      2700      1      2.727
## 3064    0141706914 4.345 2003      2700      1      2.960
## 3072    0042861328 4.269 2003      2700      1      2.818
## 3158    0041742400 4.206 2003      2700      1      2.675
## 3172    0037685217 4.392 2003      2700      1      2.861
## 3209    0038137620 4.046 2003      2700      1      2.661
## 3212    0038399801 4.509 2003      2700      1      3.124
## 3236    0037464530 4.392 2003      2700      1      3.007
```

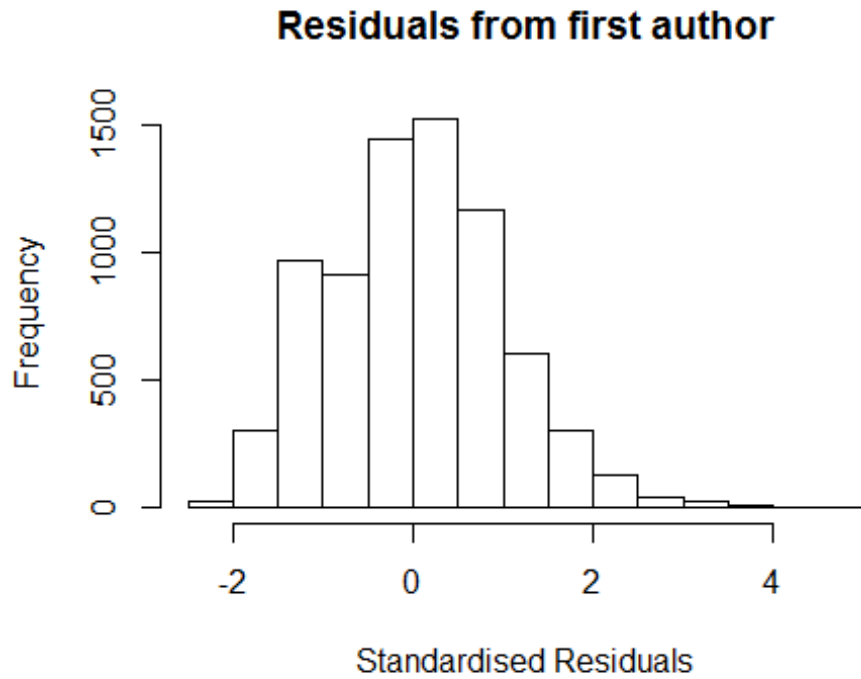
```

## 3320    0037460697 4.526 2003    2700    1    2.995
## 3375    0037355144 4.153 2003    2700    1    2.768
## 3765    2942709630 4.228 2004    2700    1    2.706
## 4001    0842265438 4.398 2004    2700    1    2.729
## 4447    20444443626 4.732 2005    2700    1    3.344
## 4449    20444442419 5.269 2005    2700    1    3.735
## 4679    11244344630 4.247 2005    2700    1    2.793
## 5375    33646114512 4.190 2006    2700    1    2.593
## 5392    33645979354 4.100 2006    2700    1    2.650
## 5438    33645052177 4.529 2006    2700    1    3.079
## 5443    33645077861 4.460 2006    2700    1    3.010
## 5524    309444466084 4.625 2006    2700    1    3.175
## 5536    31444431900 4.341 2006    2700    1    2.744
## 5540    314444450626 4.185 2006    2700    1    2.815
## 5879    34948835806 4.859 2007    2700    1    3.268
## 5883    35648953299 4.247 2007    2700    1    2.803
## 6016    34548294922 4.772 2007    2700    1    3.408
## 6022    34548047185 5.239 2007    2700    1    3.795
## 6108    34347323902 5.512 2007    2700    1    4.148
## 6182    34748875343 4.126 2007    2700    1    2.616
## 6287    34248140087 4.289 2007    2700    1    2.845
## 6299    34147179564 4.185 2007    2700    1    2.821
## 6568    33846973361 4.399 2007    2700    1    2.889
## 6627    33846543092 4.549 2007    2700    1    2.958
## 6731    58049198196 4.104 2008    2700    1    2.639
## 7188    47649111041 3.982 2008    2700    1    2.517
## 7318    48049093067 4.131 2008    2700    1    2.519
## 7529    41649095977 4.014 2008    2700    1    2.549
## 7736    38349136266 5.093 2008    2700    1    3.708
## 7862    72449172862 4.437 2009    2700    1    3.097
## 8093    70350547775 4.078 2009    2700    1    2.818
## 8283    68249144371 4.965 2009    2700    1    3.558
## 8716    63249128249 5.883 2009    2700    1    4.543
## 9175    78650205262 3.972 2010    2700    1    2.556
## 9949    77952306239 4.487 2010    2700    1    3.071
## 10035   77951046282 4.541 2010    2700    1    3.125
## 10752   80054955347 4.759 2011    2700    1    3.343
## 11023   79961054809 3.932 2011    2700    1    2.516
## 11201   79959355582 4.095 2011    2700    1    2.759
## 12893   84860702705 4.347 2012    2700    1    3.195
## 13118   84858685702 4.059 2012    2700    1    2.827
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1233 -0.6563  0.0261  0.6444  4.5426

```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.97663    0.11028   17.92 < 2e-16 ***
## FirstAuthorFemale1 0.14662    0.02425    6.05 1.6e-09 ***
## LastAuthorFemale1 -0.08033    0.02552   -3.15 0.00165 **
## Year1997        -0.00141    0.13958   -0.01 0.99194
## Year1998         0.06663    0.13128    0.51 0.61180
## Year1999        -0.04117    0.13937   -0.30 0.76768
## Year2000        -0.43736    0.26962   -1.62 0.10481
## Year2001        -0.44456    0.18063   -2.46 0.01387 *
## Year2002        -0.38738    0.12788   -3.03 0.00246 **
## Year2003        -0.59191    0.13132   -4.51 6.7e-06 ***
## Year2004        -0.45429    0.12429   -3.66 0.00026 ***
## Year2005        -0.58881    0.12361   -4.76 1.9e-06 ***
## Year2006        -0.52643    0.12075   -4.36 1.3e-05 ***
## Year2007        -0.53264    0.11699   -4.55 5.4e-06 ***
## Year2008        -0.51151    0.11577   -4.42 1.0e-05 ***
## Year2009        -0.63622    0.11563   -5.50 3.9e-08 ***
## Year2010        -0.70703    0.11354   -6.23 5.0e-10 ***
## Year2011        -0.70705    0.11527   -6.13 9.0e-10 ***
## Year2012        -0.74444    0.11295   -6.59 4.7e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.976
## Multiple R-squared:  0.0468, Adjusted R-squared:  0.0444
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 627 weights are ~= 1. The remaining 6830 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8520 0.9500 0.9090 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           1.34e-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	1.018	1	1.009
##	Year	1.018	16	1.001



```
## [1] "List of 66 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2      0029957675 5.002 1996      2700      1      2.879
## 103     0029842750 4.652 1996      2700      1      2.675
## 262     0030032146 5.242 1996      2700      1      3.346
## 471     0030764023 4.935 1997      2700      1      2.960
## 493     0030835927 4.545 1997      2700      1      2.503
## 536     0030859363 4.598 1997      2700      1      2.623
## 636     0031041139 5.114 1997      2700      1      3.219
## 821     0032512287 4.574 1998      2700      1      2.531
## 1590    0033585476 5.332 1999      2700      1      3.397
## 1597    0033573788 4.886 1999      2700      1      2.804
## 2135    0035292711 4.056 2001      2700      1      2.604
## 2332    0035661067 4.269 2001      2700      1      2.737
## 2705    0037070806 4.592 2002      2700      1      3.083
## 2738    0037035126 4.738 2002      2700      1      3.082
## 2764    0037148929 5.770 2002      2700      1      4.261
## 2773    0037116811 4.433 2002      2700      1      2.844
## 2940    0344666767 5.279 2003      2700      1      3.894
## 2991    0345374593 4.258 2003      2700      1      2.727
## 3064    0141706914 4.345 2003      2700      1      2.960
## 3072    0042861328 4.269 2003      2700      1      2.818
## 3158    0041742400 4.206 2003      2700      1      2.675
```

```

## 3172    0037685217 4.392 2003    2700    1    2.861
## 3209    0038137620 4.046 2003    2700    1    2.661
## 3212    0038399801 4.509 2003    2700    1    3.124
## 3236    0037464530 4.392 2003    2700    1    3.007
## 3320    0037460697 4.526 2003    2700    1    2.995
## 3375    0037355144 4.153 2003    2700    1    2.768
## 3765    2942709630 4.228 2004    2700    1    2.706
## 4001    0842265438 4.398 2004    2700    1    2.729
## 4447    20444443626 4.732 2005    2700    1    3.344
## 4449    20444442419 5.269 2005    2700    1    3.735
## 4679    11244344630 4.247 2005    2700    1    2.793
## 5375    33646114512 4.190 2006    2700    1    2.593
## 5392    33645979354 4.100 2006    2700    1    2.650
## 5438    33645052177 4.529 2006    2700    1    3.079
## 5443    33645077861 4.460 2006    2700    1    3.010
## 5524    309444466084 4.625 2006    2700    1    3.175
## 5536    31444431900 4.341 2006    2700    1    2.744
## 5540    314444450626 4.185 2006    2700    1    2.815
## 5879    34948835806 4.859 2007    2700    1    3.268
## 5883    35648953299 4.247 2007    2700    1    2.803
## 6016    34548294922 4.772 2007    2700    1    3.408
## 6022    34548047185 5.239 2007    2700    1    3.795
## 6108    34347323902 5.512 2007    2700    1    4.148
## 6182    34748875343 4.126 2007    2700    1    2.616
## 6287    34248140087 4.289 2007    2700    1    2.845
## 6299    34147179564 4.185 2007    2700    1    2.821
## 6568    33846973361 4.399 2007    2700    1    2.889
## 6627    33846543092 4.549 2007    2700    1    2.958
## 6731    58049198196 4.104 2008    2700    1    2.639
## 7188    47649111041 3.982 2008    2700    1    2.517
## 7318    48049093067 4.131 2008    2700    1    2.519
## 7529    41649095977 4.014 2008    2700    1    2.549
## 7736    38349136266 5.093 2008    2700    1    3.708
## 7862    72449172862 4.437 2009    2700    1    3.097
## 8093    70350547775 4.078 2009    2700    1    2.818
## 8283    68249144371 4.965 2009    2700    1    3.558
## 8716    63249128249 5.883 2009    2700    1    4.543
## 9175    78650205262 3.972 2010    2700    1    2.556
## 9949    77952306239 4.487 2010    2700    1    3.071
## 10035   77951046282 4.541 2010    2700    1    3.125
## 10752   80054955347 4.759 2011    2700    1    3.343
## 11023   79961054809 3.932 2011    2700    1    2.516
## 11201   79959355582 4.095 2011    2700    1    2.759
## 12893   84860702705 4.347 2012    2700    1    3.195
## 13118   84858685702 4.059 2012    2700    1    2.827

```

```
##
```

```
## Call:
```

```

## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))

```

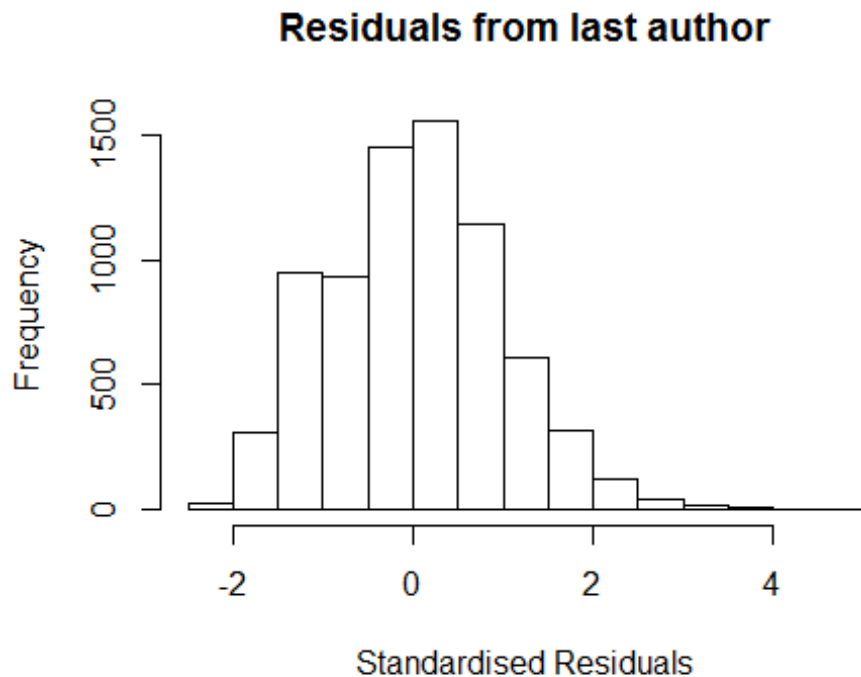


```

## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1534 -0.6665  0.0336  0.6451  4.5567
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.96507    0.10990   17.88 < 2e-16 ***
## FirstAuthorFemale1 0.11862    0.02339    5.07 4.0e-07 ***
## Year1997        0.00316    0.13899    0.02 0.98189
## Year1998        0.06968    0.13119    0.53 0.59535
## Year1999       -0.04367    0.13898   -0.31 0.75338
## Year2000       -0.45163    0.26798   -1.69 0.09197 .
## Year2001       -0.45250    0.17987   -2.52 0.01190 *
## Year2002       -0.39013    0.12765   -3.06 0.00225 **
## Year2003       -0.59740    0.13095   -4.56 5.2e-06 ***
## Year2004       -0.45616    0.12398   -3.68 0.00024 ***
## Year2005       -0.59207    0.12326   -4.80 1.6e-06 ***
## Year2006       -0.53067    0.12048   -4.40 1.1e-05 ***
## Year2007       -0.53752    0.11669   -4.61 4.2e-06 ***
## Year2008       -0.51865    0.11544   -4.49 7.1e-06 ***
## Year2009       -0.63876    0.11535   -5.54 3.2e-08 ***
## Year2010       -0.70929    0.11326   -6.26 4.0e-10 ***
## Year2011       -0.70941    0.11503   -6.17 7.3e-10 ***
## Year2012       -0.74810    0.11268   -6.64 3.4e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.978
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0433
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 607 weights are ~= 1. The remaining 6850 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8550 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          1.34e-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 1.021 1          1.010
## Year             1.021 16          1.001
```



```
## [1] "List of 66 outliers with residuals above 2.5"
##          ScopusId  NLCS Year OneField Fields residuals
## 2          0029957675 5.002 1996      2700      1      2.879
## 103         0029842750 4.652 1996      2700      1      2.675
## 262         0030032146 5.242 1996      2700      1      3.346
## 471         0030764023 4.935 1997      2700      1      2.960
## 493         0030835927 4.545 1997      2700      1      2.503
## 536         0030859363 4.598 1997      2700      1      2.623
## 636         0031041139 5.114 1997      2700      1      3.219
## 821         0032512287 4.574 1998      2700      1      2.531
## 1590        0033585476 5.332 1999      2700      1      3.397
## 1597        0033573788 4.886 1999      2700      1      2.804
## 2135        0035292711 4.056 2001      2700      1      2.604
## 2332        0035661067 4.269 2001      2700      1      2.737
## 2705        0037070806 4.592 2002      2700      1      3.083
## 2738        0037035126 4.738 2002      2700      1      3.082
## 2764        0037148929 5.770 2002      2700      1      4.261
## 2773        0037116811 4.433 2002      2700      1      2.844
## 2940        0344666767 5.279 2003      2700      1      3.894
## 2991        0345374593 4.258 2003      2700      1      2.727
```

## 3064	0141706914	4.345	2003	2700	1	2.960
## 3072	0042861328	4.269	2003	2700	1	2.818
## 3158	0041742400	4.206	2003	2700	1	2.675
## 3172	0037685217	4.392	2003	2700	1	2.861
## 3209	0038137620	4.046	2003	2700	1	2.661
## 3212	0038399801	4.509	2003	2700	1	3.124
## 3236	0037464530	4.392	2003	2700	1	3.007
## 3320	0037460697	4.526	2003	2700	1	2.995
## 3375	0037355144	4.153	2003	2700	1	2.768
## 3765	2942709630	4.228	2004	2700	1	2.706
## 4001	0842265438	4.398	2004	2700	1	2.729
## 4447	204444443626	4.732	2005	2700	1	3.344
## 4449	204444442419	5.269	2005	2700	1	3.735
## 4679	11244344630	4.247	2005	2700	1	2.793
## 5375	33646114512	4.190	2006	2700	1	2.593
## 5392	33645979354	4.100	2006	2700	1	2.650
## 5438	33645052177	4.529	2006	2700	1	3.079
## 5443	33645077861	4.460	2006	2700	1	3.010
## 5524	309444466084	4.625	2006	2700	1	3.175
## 5536	31444431900	4.341	2006	2700	1	2.744
## 5540	31444450626	4.185	2006	2700	1	2.815
## 5879	34948835806	4.859	2007	2700	1	3.268
## 5883	35648953299	4.247	2007	2700	1	2.803
## 6016	34548294922	4.772	2007	2700	1	3.408
## 6022	34548047185	5.239	2007	2700	1	3.795
## 6108	34347323902	5.512	2007	2700	1	4.148
## 6182	34748875343	4.126	2007	2700	1	2.616
## 6287	34248140087	4.289	2007	2700	1	2.845
## 6299	34147179564	4.185	2007	2700	1	2.821
## 6568	33846973361	4.399	2007	2700	1	2.889
## 6627	33846543092	4.549	2007	2700	1	2.958
## 6731	58049198196	4.104	2008	2700	1	2.639
## 7188	47649111041	3.982	2008	2700	1	2.517
## 7318	48049093067	4.131	2008	2700	1	2.519
## 7529	41649095977	4.014	2008	2700	1	2.549
## 7736	38349136266	5.093	2008	2700	1	3.708
## 7862	72449172862	4.437	2009	2700	1	3.097
## 8093	70350547775	4.078	2009	2700	1	2.818
## 8283	68249144371	4.965	2009	2700	1	3.558
## 8716	63249128249	5.883	2009	2700	1	4.543
## 9175	78650205262	3.972	2010	2700	1	2.556
## 9949	77952306239	4.487	2010	2700	1	3.071
## 10035	77951046282	4.541	2010	2700	1	3.125
## 10752	80054955347	4.759	2011	2700	1	3.343
## 11023	79961054809	3.932	2011	2700	1	2.516
## 11201	79959355582	4.095	2011	2700	1	2.759
## 12893	84860702705	4.347	2012	2700	1	3.195
## 13118	84858685702	4.059	2012	2700	1	2.827

##

Call:

```

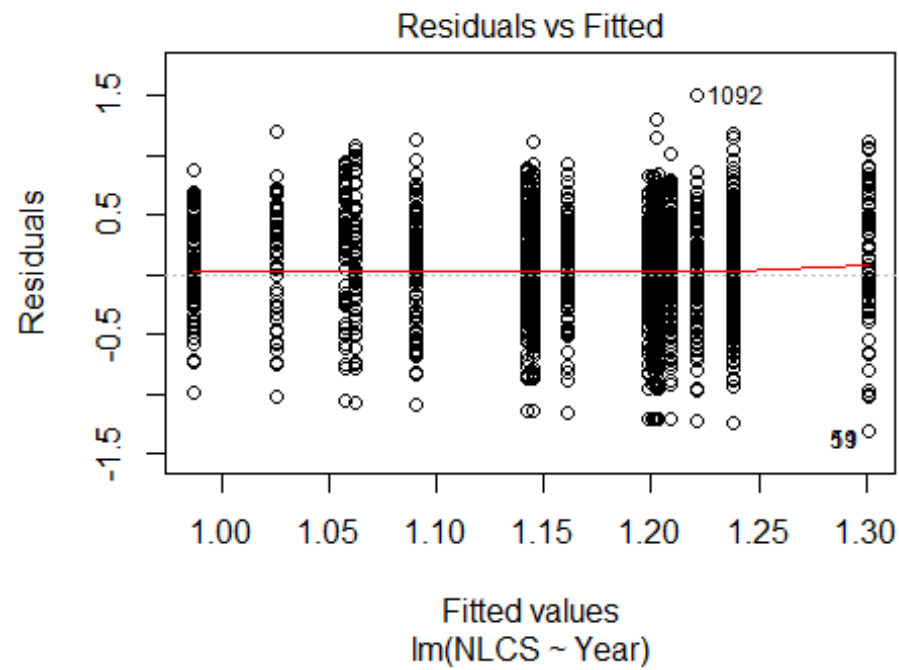
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0845 -0.6538  0.0294  0.6517  4.4897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.01173    0.10989   18.31 < 2e-16 ***
## LastAuthorFemale1 -0.02438    0.02440   -1.00  0.31766
## Year1997          0.00272    0.13950    0.02  0.98443
## Year1998          0.07276    0.13135    0.55  0.57966
## Year1999         -0.04092    0.13921   -0.29  0.76882
## Year2000         -0.45237    0.26709   -1.69  0.09036 .
## Year2001         -0.42264    0.18059   -2.34  0.01929 *
## Year2002         -0.37991    0.12766   -2.98  0.00293 **
## Year2003         -0.58798    0.13126   -4.48  7.6e-06 ***
## Year2004         -0.44808    0.12412   -3.61  0.00031 ***
## Year2005         -0.58361    0.12347   -4.73  2.3e-06 ***
## Year2006         -0.51428    0.12054   -4.27  2.0e-05 ***
## Year2007         -0.52097    0.11673   -4.46  8.2e-06 ***
## Year2008         -0.49740    0.11551   -4.31  1.7e-05 ***
## Year2009         -0.61843    0.11533   -5.36  8.5e-08 ***
## Year2010         -0.69106    0.11327   -6.10  1.1e-09 ***
## Year2011         -0.68817    0.11497   -5.99  2.3e-09 ***
## Year2012         -0.73078    0.11268   -6.49  9.4e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.979
## Multiple R-squared:  0.0422, Adjusted R-squared:  0.04
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 624 weights are ~ = 1. The remaining 6833 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0018 0.8480 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          1.34e-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

```

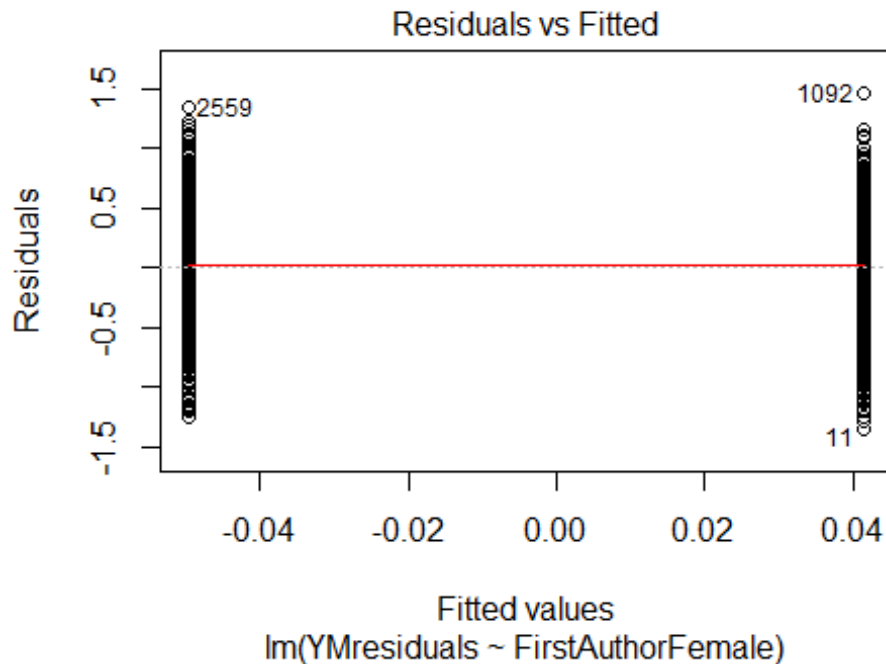
```

##                psi                subsampling                cov
##                "bisquare"          "nonsingular"            ".vcov.avar1"
## compute.outlier.stats
##                "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7457"
## [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
## 142 131 101 108 150 228 166 147 138 150 224 248 287 286 276
## 2011 2012
## 330 395
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 85 72 65 88 95 116 108 103 121 175 171 209 216 211
## 2011 2012
## 229 267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 76 65 60 83 86 111 92 80 101 152 149 190 193 193
## 2011 2012
## 213 246
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 84, df = 16, p-value = 3e-11

```

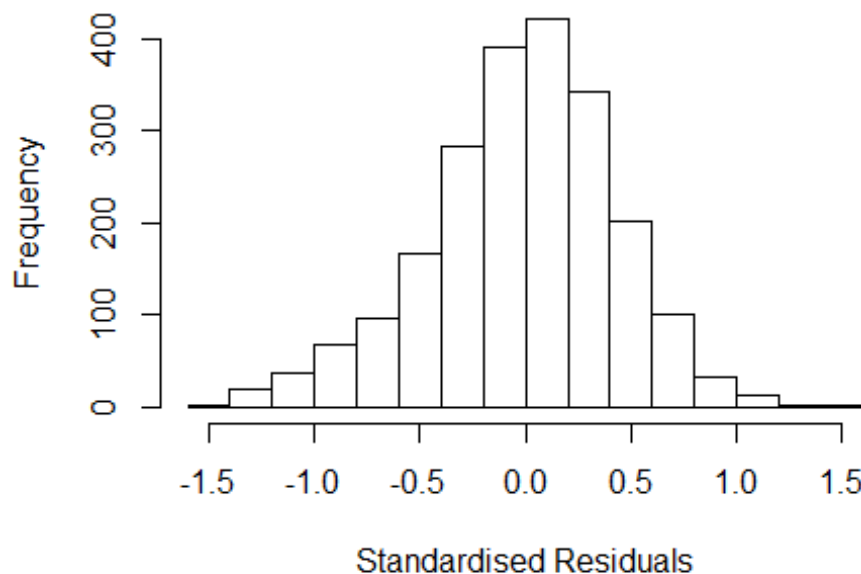


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 23, df = 1, p-value = 2e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.77216675748442"
## [1] "Male first author team size 2018 geometric mean: 3.92213939089852"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.33759694180786"
## [1] "Male last author team size 2018 geometric mean: 4.5958160541434"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.108 1      1.053
## LastAuthorFemale  1.072 1      1.035
## UniqueAuthors    1.313 4      1.035
## Year              1.419 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
## -1.4153 -0.2749  0.0124  0.2797  1.4438
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14053    0.07376   15.46 < 2e-16 ***
## FirstAuthorFemale1  0.05608    0.02012    2.79  0.00536 **
## LastAuthorFemale1 -0.00275    0.01977   -0.14  0.88938
## UniqueAuthors2     0.21868    0.04422    4.95  8.2e-07 ***
## UniqueAuthors3     0.30938    0.04164    7.43  1.6e-13 ***
## UniqueAuthors4     0.33675    0.04222    7.98  2.4e-15 ***
## UniqueAuthors5     0.43907    0.04008   10.96 < 2e-16 ***
## Year1997          -0.34038    0.10828   -3.14  0.00169 **
## Year1998          -0.33268    0.10296   -3.23  0.00125 **
## Year1999          -0.33251    0.09311   -3.57  0.00036 ***
```

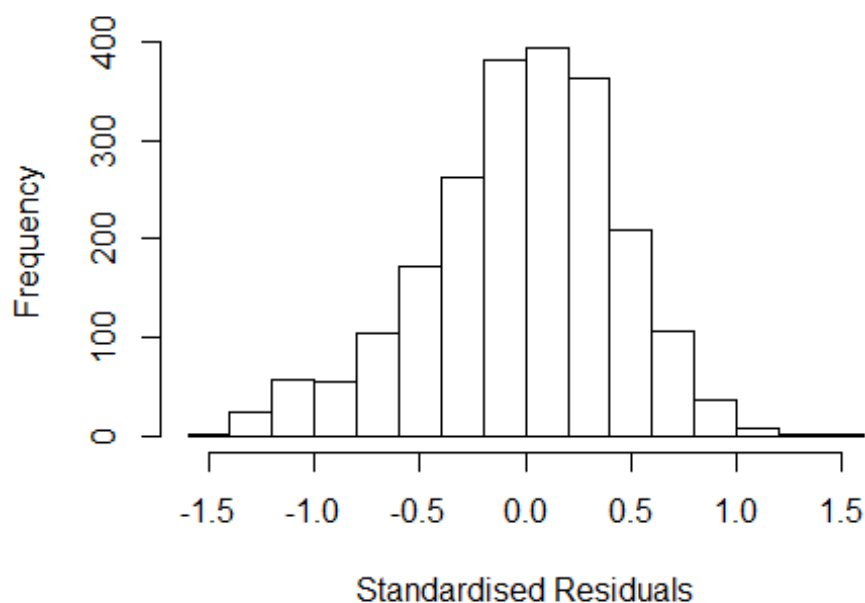


```

## Year2000      -0.38122    0.08114   -4.70  2.8e-06 ***
## Year2001      -0.26374    0.08977   -2.94  0.00334 **
## Year2002      -0.22499    0.07750   -2.90  0.00373 **
## Year2003      -0.23088    0.07668   -3.01  0.00264 **
## Year2004      -0.27590    0.08166   -3.38  0.00074 ***
## Year2005      -0.26990    0.08125   -3.32  0.00091 ***
## Year2006      -0.25495    0.07401   -3.44  0.00058 ***
## Year2007      -0.27579    0.07397   -3.73  0.00020 ***
## Year2008      -0.27321    0.07432   -3.68  0.00024 ***
## Year2009      -0.30431    0.07311   -4.16  3.3e-05 ***
## Year2010      -0.34936    0.07287   -4.79  1.7e-06 ***
## Year2011      -0.28718    0.07262   -3.95  7.9e-05 ***
## Year2012      -0.22805    0.07323   -3.11  0.00187 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.119, Adjusted R-squared:  0.11
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 2005 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.863  0.953  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      4.59e-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 1.073 1      1.036
## LastAuthorFemale  1.058 1      1.028
## Year              1.098 16      1.003

```

Residuals from first and last author



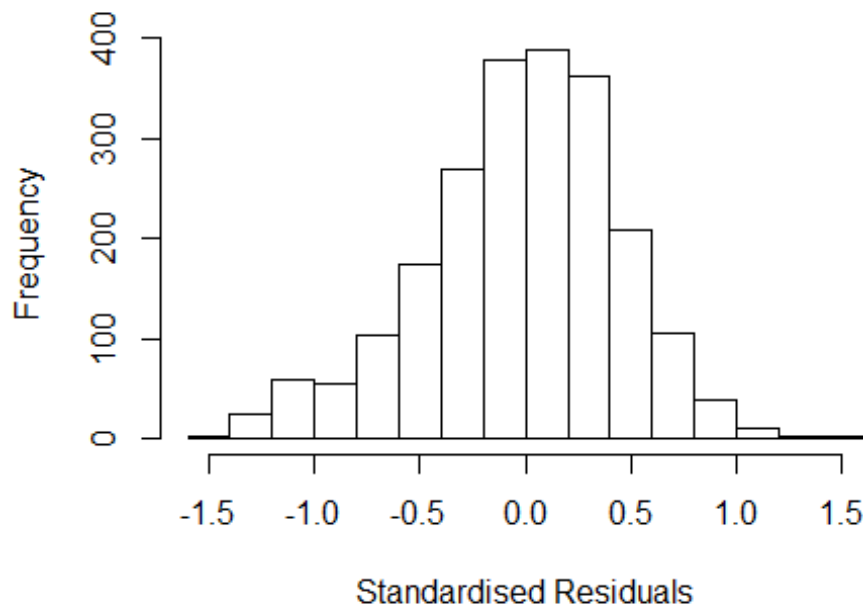
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45848 -0.28181  0.00964  0.29286  1.44457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3712     0.0741   18.50 < 2e-16 ***
## FirstAuthorFemale1  0.0873     0.0208    4.20 2.8e-05 ***
## LastAuthorFemale1 -0.0180     0.0207   -0.87 0.38435
## Year1997          -0.3537     0.1165   -3.04 0.00243 **
## Year1998          -0.3214     0.1129   -2.85 0.00445 **
## Year1999          -0.3499     0.1094   -3.20 0.00140 **
## Year2000          -0.3956     0.0900   -4.39 1.2e-05 ***
## Year2001          -0.2651     0.0956   -2.77 0.00558 **
## Year2002          -0.1631     0.0844   -1.93 0.05356 .
## Year2003          -0.1751     0.0829   -2.11 0.03480 *
## Year2004          -0.1999     0.0864   -2.31 0.02075 *
## Year2005          -0.2243     0.0870   -2.58 0.01005 *
```

```

## Year2006          -0.1981      0.0801   -2.47   0.01350 *
## Year2007          -0.2006      0.0806   -2.49   0.01284 *
## Year2008          -0.2038      0.0799   -2.55   0.01085 *
## Year2009          -0.2449      0.0797   -3.07   0.00216 **
## Year2010          -0.2633      0.0797   -3.30   0.00097 ***
## Year2011          -0.1868      0.0784   -2.38   0.01723 *
## Year2012          -0.1424      0.0794   -1.79   0.07301 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0382, Adjusted R-squared:  0.0302
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 186 weights are ~= 1. The remaining 1993 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.224  0.862  0.950  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      4.59e-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 1.057 1      1.028
## Year      1.057 16      1.002

```

Residuals from first author



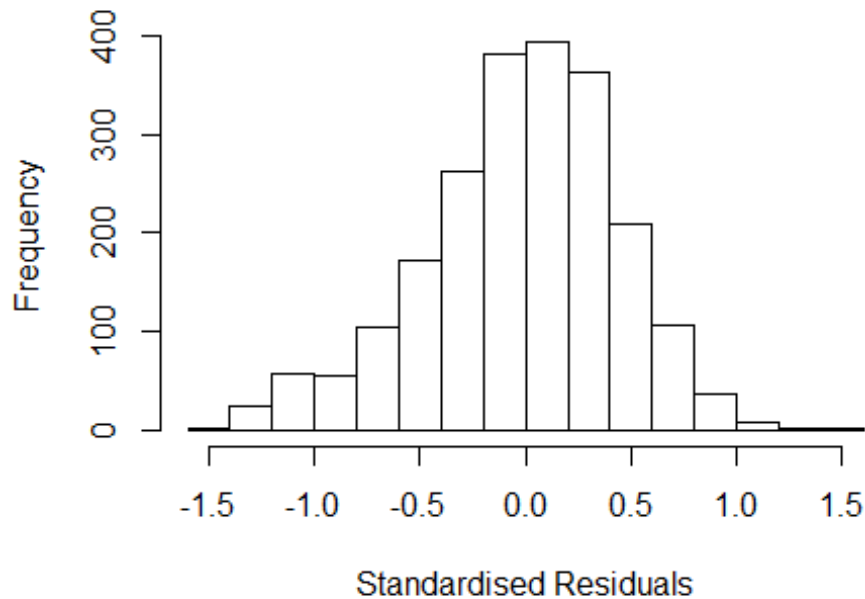
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4513 -0.2860  0.0101  0.2949  1.4332
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3669    0.0739   18.50 < 2e-16 ***
## FirstAuthorFemale1  0.0844    0.0206    4.09 4.5e-05 ***
## Year1997         -0.3548    0.1168   -3.04 0.00240 **
## Year1998         -0.3207    0.1132   -2.83 0.00465 **
## Year1999         -0.3514    0.1093   -3.21 0.00132 **
## Year2000         -0.3967    0.0902   -4.40 1.1e-05 ***
## Year2001         -0.2647    0.0956   -2.77 0.00567 **
## Year2002         -0.1626    0.0845   -1.92 0.05466 .
## Year2003         -0.1733    0.0829   -2.09 0.03665 *
## Year2004         -0.1990    0.0865   -2.30 0.02158 *
## Year2005         -0.2254    0.0873   -2.58 0.00983 **
## Year2006         -0.2004    0.0803   -2.50 0.01264 *
```

```

## Year2007          -0.2010      0.0808   -2.49  0.01295 *
## Year2008          -0.2053      0.0801   -2.56  0.01048 *
## Year2009          -0.2461      0.0799   -3.08  0.00210 **
## Year2010          -0.2647      0.0799   -3.31  0.00094 ***
## Year2011          -0.1880      0.0786   -2.39  0.01687 *
## Year2012          -0.1448      0.0796   -1.82  0.06908 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0379, Adjusted R-squared:  0.0304
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 191 weights are ~= 1. The remaining 1988 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.229  0.860  0.950  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      4.59e-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 1.043 1      1.021
## Year              1.043 16      1.001

```

Residuals from last author



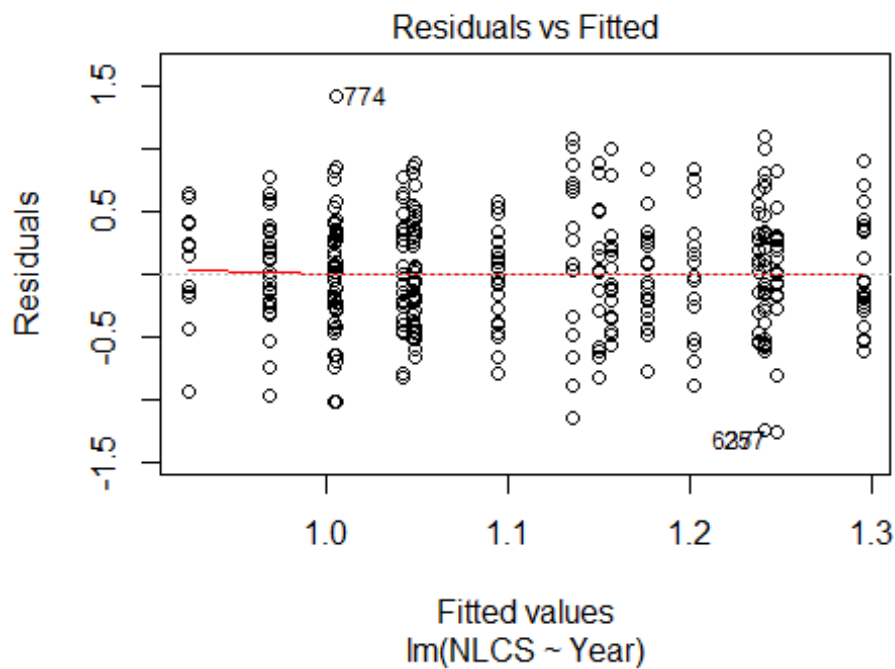
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3963 -0.2893 0.0191 0.2953 1.4822
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.39629 0.07374 18.93 < 2e-16 ***
## LastAuthorFemale1 -0.00364 0.02048 -0.18 0.8590
## Year1997 -0.35021 0.11517 -3.04 0.0024 **
## Year1998 -0.31271 0.11343 -2.76 0.0059 **
## Year1999 -0.34338 0.11027 -3.11 0.0019 **
## Year2000 -0.39525 0.08966 -4.41 1.1e-05 ***
## Year2001 -0.25090 0.09566 -2.62 0.0088 **
## Year2002 -0.15286 0.08415 -1.82 0.0695 .
## Year2003 -0.15456 0.08260 -1.87 0.0615 .
## Year2004 -0.17753 0.08628 -2.06 0.0397 *
## Year2005 -0.20609 0.08652 -2.38 0.0173 *
## Year2006 -0.18133 0.07972 -2.27 0.0230 *
```

```

## Year2007          -0.18074      0.08053    -2.24    0.0249 *
## Year2008          -0.18236      0.07982    -2.28    0.0224 *
## Year2009          -0.22033      0.07937    -2.78    0.0056 **
## Year2010          -0.23904      0.07921    -3.02    0.0026 **
## Year2011          -0.16137      0.07799    -2.07    0.0387 *
## Year2012          -0.11552      0.07874    -1.47    0.1425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.0219
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 1974 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.209  0.859  0.948  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      4.59e-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: 2179"
## [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
##   40   53   43   52   39   40   54   44   26   38   39   33   39   19   30
## 2011 2012
##   20   28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   25   24   32   17   16   29   24   20   20   24   14   27   16   17
## 2011 2012

```

```
## 13 23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 23 20 28 15 14 26 22 16 16 19 12 24 10 16
## 2011 2012
## 10 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 = 25, df = 16, p-value = 0.07
```

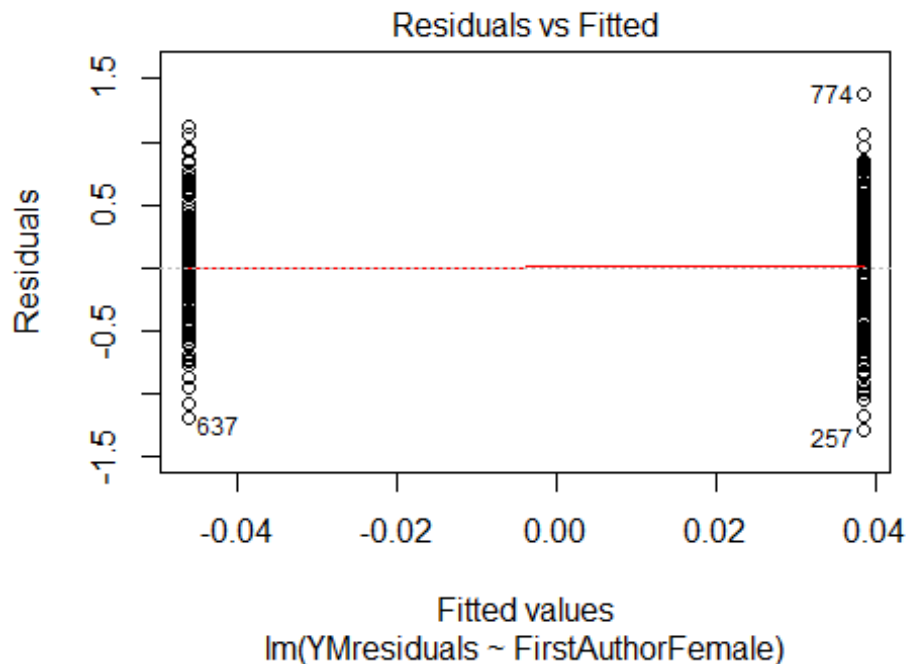


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
## [1] "Female first author team size 2018 geometric mean: 3.90157254239273"
## [1] "Male first author team size 2018 geometric mean: 3.26376288449377"
## 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 = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.53003577953455"
## [1] "Male last author team size 2018 geometric mean: 3.41532392662992"

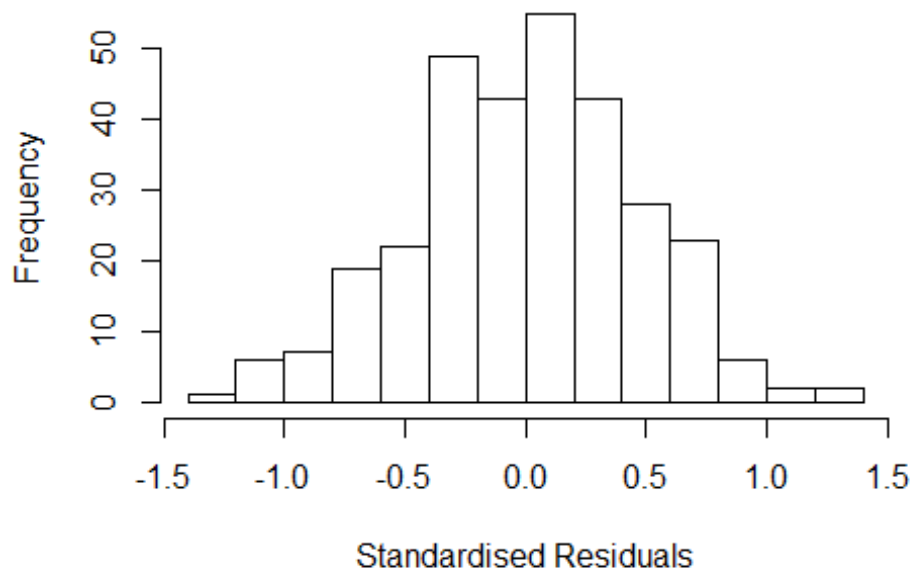
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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"
##
```

	GVIF	Df	GVIF ^{1/(2*Df)}
FirstAuthorFemale	1.483	1	1.218
LastAuthorFemale	1.392	1	1.180
UniqueAuthors	2.407	4	1.116
Year	2.919	16	1.034

Residuals from first and last author and team size



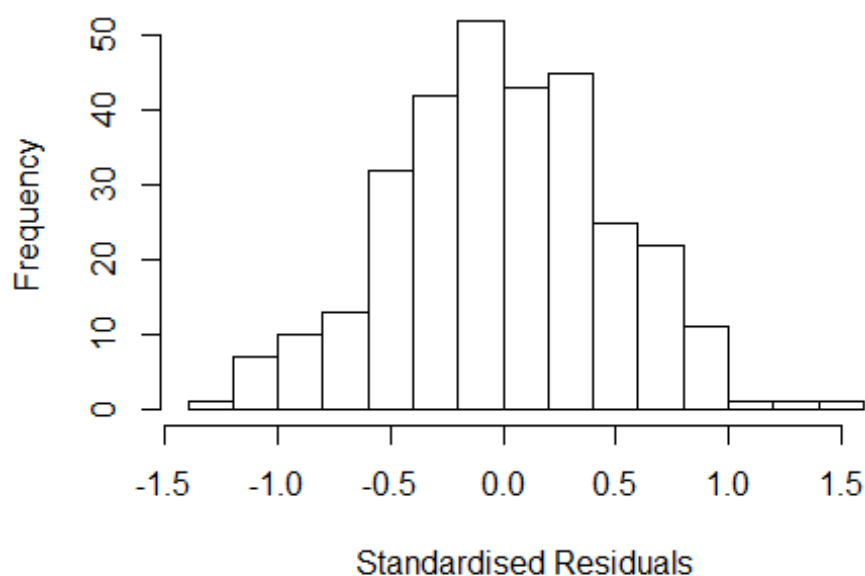
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3030 -0.3038  0.0104  0.2931  1.3248
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.87676    0.40839   2.15   0.0327 *
## FirstAuthorFemale1 -0.10823    0.06192  -1.75   0.0816 .
## LastAuthorFemale1  0.00569    0.06162   0.09   0.9266
## UniqueAuthors2    0.12793    0.08981   1.42   0.1554
## UniqueAuthors3    0.20665    0.09115   2.27   0.0241 *
## UniqueAuthors4    0.31927    0.10824   2.95   0.0034 **
## UniqueAuthors5    0.45380    0.09765   4.65 5.2e-06 ***
## Year1997         0.23023    0.40673   0.57   0.5718
## Year1998         0.08186    0.40844   0.20   0.8413
## Year1999         0.01977    0.41095   0.05   0.9617
```

```

## Year2000      0.27107      0.42423      0.64      0.5234
## Year2001      0.09853      0.41653      0.24      0.8132
## Year2002     -0.02988      0.41458     -0.07      0.9426
## Year2003      0.01489      0.41506      0.04      0.9714
## Year2004      0.11055      0.41799      0.26      0.7916
## Year2005     -0.03966      0.41661     -0.10      0.9242
## Year2006     -0.04219      0.41853     -0.10      0.9198
## Year2007      0.00870      0.43355      0.02      0.9840
## Year2008      0.08061      0.42273      0.19      0.8489
## Year2009      0.11733      0.43754      0.27      0.7888
## Year2010      0.17916      0.41369      0.43      0.6653
## Year2011     -0.12584      0.44544     -0.28      0.7778
## Year2012     -0.10651      0.43627     -0.24      0.8073
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.141, Adjusted R-squared:  0.0747
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.386  0.869  0.955   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      3.27e-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 1.426 1      1.194
## LastAuthorFemale  1.195 1      1.093
## Year              1.443 16      1.012

```

Residuals from first and last author



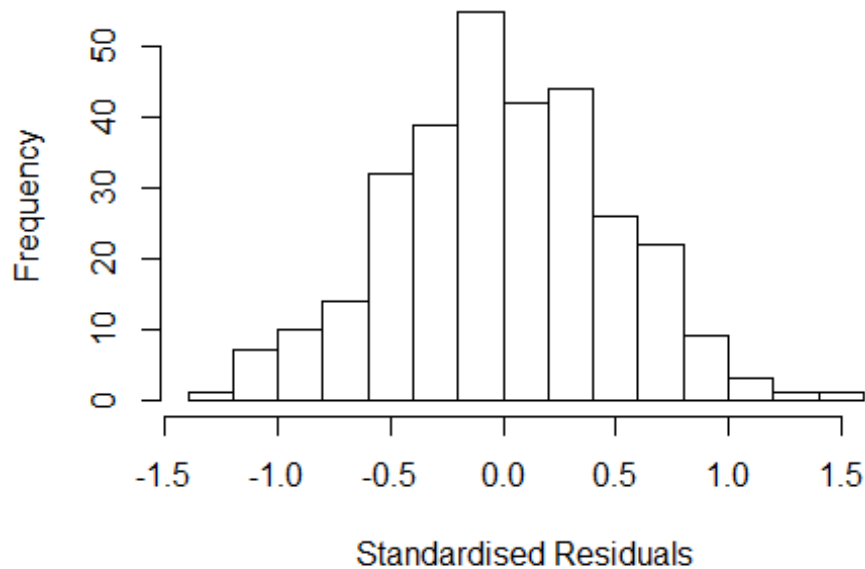
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.321 -0.319 -0.011 0.304 1.427
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1091 0.3599 3.08 0.0023 **
## FirstAuthorFemale1 -0.1122 0.0649 -1.73 0.0848 .
## LastAuthorFemale1 -0.0286 0.0620 -0.46 0.6449
## Year1997 0.2047 0.3707 0.55 0.5812
## Year1998 0.0262 0.3753 0.07 0.9445
## Year1999 -0.0288 0.3735 -0.08 0.9386
## Year2000 0.2116 0.3854 0.55 0.5833
## Year2001 0.1299 0.3862 0.34 0.7369
## Year2002 -0.0887 0.3752 -0.24 0.8133
## Year2003 0.0230 0.3783 0.06 0.9515
## Year2004 0.1617 0.3780 0.43 0.6691
## Year2005 0.0298 0.3809 0.08 0.9378
```

```

## Year2006          0.0108      0.3765      0.03      0.9770
## Year2007          0.0571      0.4055      0.14      0.8881
## Year2008          0.1410      0.3840      0.37      0.7138
## Year2009          0.1959      0.3976      0.49      0.6227
## Year2010          0.1840      0.3789      0.49      0.6275
## Year2011         -0.0897      0.4238     -0.21      0.8326
## Year2012         -0.0870      0.3975     -0.22      0.8269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0585, Adjusted R-squared:  -0.000516
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 282 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.402  0.886  0.958  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      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 1.332 1      1.154
## Year              1.332 16      1.009

```

Residuals from first author



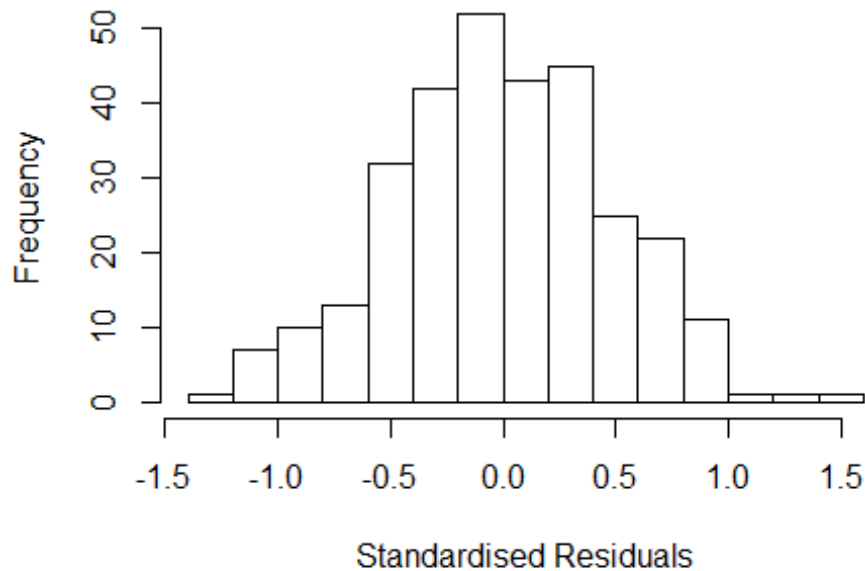
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3142 -0.3130 -0.0238 0.3108 1.4043
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10481 0.36516 3.03 0.0027 **
## FirstAuthorFemale1 -0.11901 0.06275 -1.90 0.0589 .
## Year1997 0.20153 0.37740 0.53 0.5938
## Year1998 0.02129 0.38260 0.06 0.9557
## Year1999 -0.02999 0.38026 -0.08 0.9372
## Year2000 0.20941 0.39240 0.53 0.5940
## Year2001 0.13275 0.39197 0.34 0.7351
## Year2002 -0.08775 0.38154 -0.23 0.8183
## Year2003 0.02190 0.38536 0.06 0.9547
## Year2004 0.16085 0.38499 0.42 0.6764
## Year2005 0.02933 0.38770 0.08 0.9397
## Year2006 0.00799 0.38435 0.02 0.9834
```

```

## Year2007          0.05555    0.41216    0.13    0.8929
## Year2008          0.13650    0.39174    0.35    0.7278
## Year2009          0.19934    0.40303    0.49    0.6213
## Year2010          0.18873    0.38416    0.49    0.6236
## Year2011         -0.08553    0.42901   -0.20    0.8421
## Year2012         -0.08906    0.40375   -0.22    0.8256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.058, Adjusted R-squared:  0.00244
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.412  0.880  0.957  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      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 1.111 1          1.054
## Year            1.111 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
## -1.27776 -0.31027 -0.00369 0.33191 1.49646
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08800 0.33593 3.24 0.0013 **
## LastAuthorFemale1 -0.05871 0.06015 -0.98 0.3299
## Year1997 0.20470 0.34656 0.59 0.5552
## Year1998 0.01131 0.34799 0.03 0.9741
## Year1999 -0.05045 0.34550 -0.15 0.8840
## Year2000 0.18976 0.35587 0.53 0.5943
## Year2001 0.09536 0.35566 0.27 0.7888
## Year2002 -0.11239 0.34708 -0.32 0.7463
## Year2003 -0.00447 0.34955 -0.01 0.9898
## Year2004 0.12662 0.34969 0.36 0.7175
## Year2005 0.00578 0.35307 0.02 0.9870
## Year2006 -0.02867 0.34690 -0.08 0.9342
```

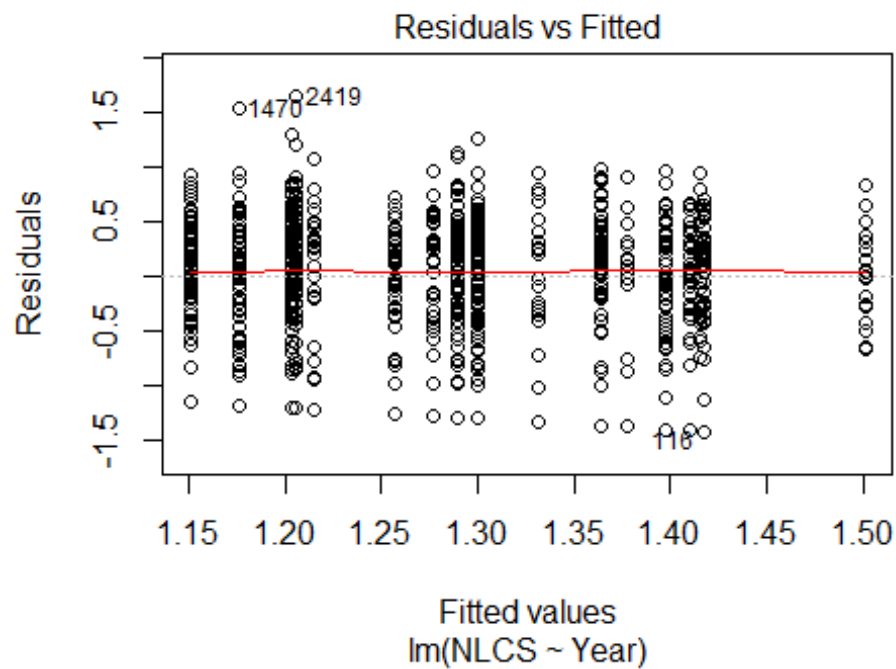


```

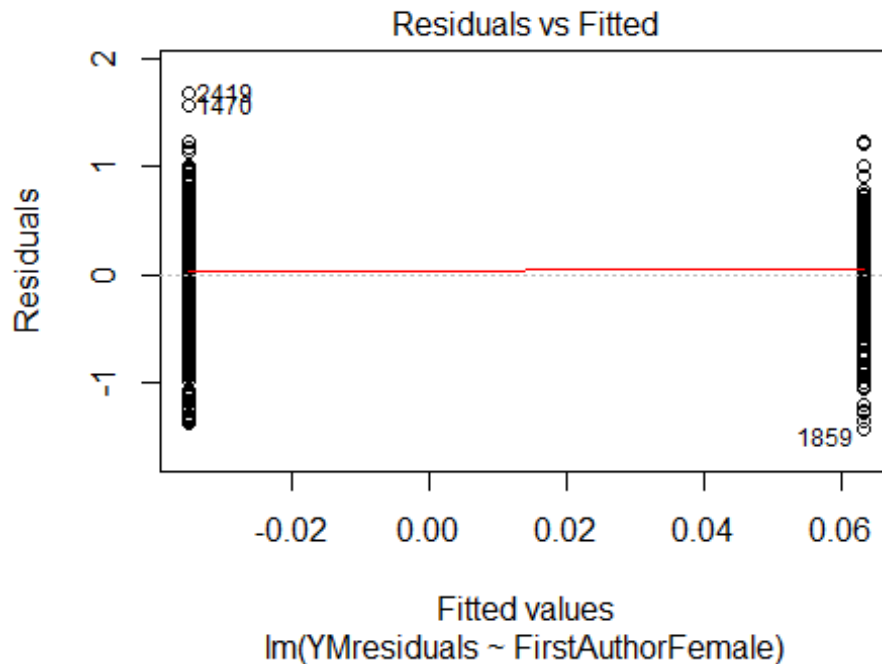
## Year2007      0.04522      0.38772      0.12      0.9072
## Year2008      0.09845      0.35434      0.28      0.7813
## Year2009      0.17391      0.37364      0.47      0.6420
## Year2010      0.16623      0.35403      0.47      0.6390
## Year2011     -0.15839      0.39062     -0.41      0.6854
## Year2012     -0.10575      0.37397     -0.28      0.7776
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.0477, Adjusted R-squared:  -0.00852
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 285 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.358  0.885   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      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 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
##   92  116   91   96  100   96  103   80  104  114  154  154  132  162  164
## 2011 2012
##  185  176
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   42   19   24   30   28   33   29   44   50   91   66   86   86   75
## 2011 2012

```

```
## 112 96
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 39 17 24 28 23 28 23 36 46 64 59 72 68 70
## 2011 2012
## 103 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 = 16, df = 16, p-value = 0.4
```

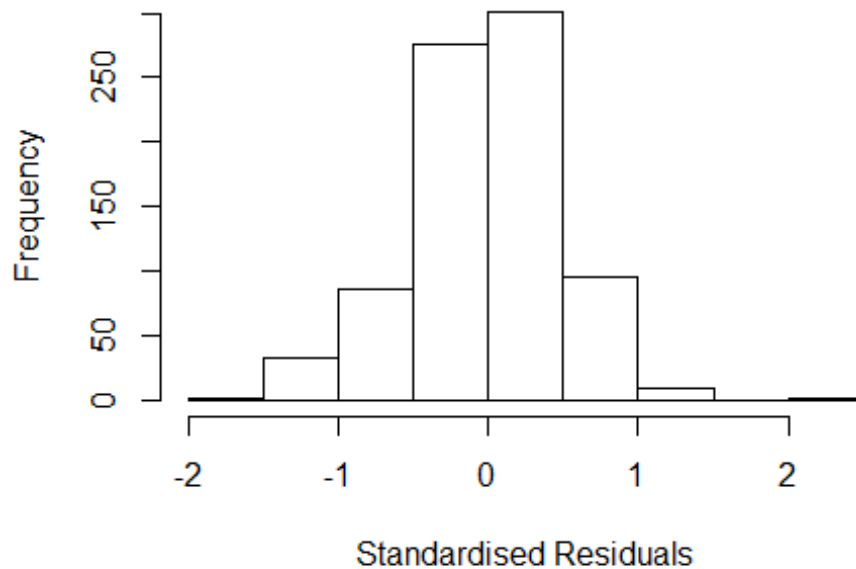


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.8, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 4.33941325796633"
## [1] "Male first author team size 2018 geometric mean: 3.45961403855832"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.8764020673988"
## [1] "Male last author team size 2018 geometric mean: 3.89405095985371"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, 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 1.149  1      1.072
## LastAuthorFemale  1.071  1      1.035
## UniqueAuthors    1.649  4      1.065
## Year             1.801 16      1.019
```

Residuals from first and last author and team size



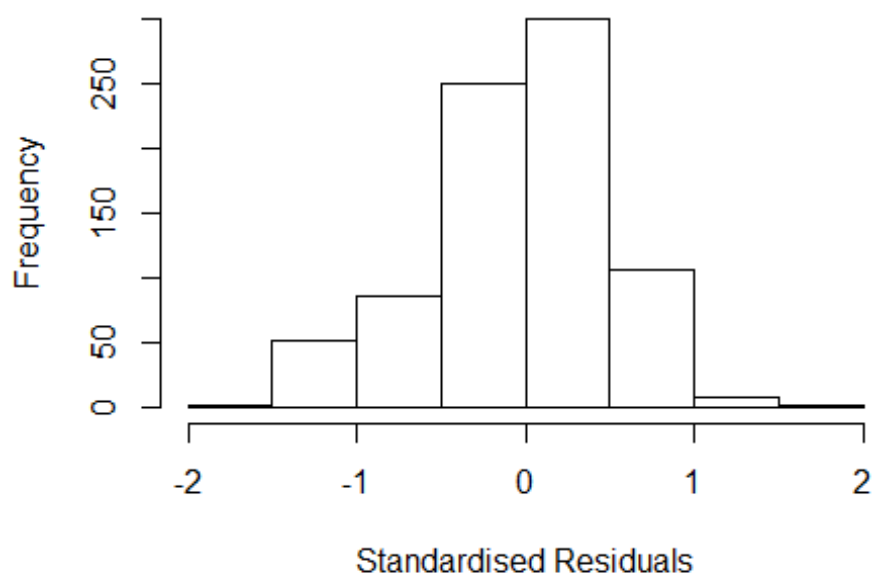
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.53835 -0.31333 0.00762 0.31029 2.17984
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8846 0.1408 6.28 5.6e-10 ***
## FirstAuthorFemale1 0.1112 0.0360 3.09 0.0021 **
## LastAuthorFemale1 0.0578 0.0369 1.57 0.1175
## UniqueAuthors2 0.4804 0.0931 5.16 3.2e-07 ***
## UniqueAuthors3 0.4836 0.0916 5.28 1.7e-07 ***
## UniqueAuthors4 0.5895 0.0913 6.45 1.9e-10 ***
## UniqueAuthors5 0.7505 0.0862 8.70 < 2e-16 ***
## Year1997 0.0237 0.1472 0.16 0.8721
## Year1998 0.0503 0.1511 0.33 0.7392
## Year1999 -0.0551 0.1795 -0.31 0.7589
```

```

## Year2000          -0.1745      0.1630    -1.07    0.2847
## Year2001          -0.2990      0.1577    -1.90    0.0584 .
## Year2002          -0.0624      0.1481    -0.42    0.6737
## Year2003          -0.1331      0.1463    -0.91    0.3634
## Year2004          -0.1549      0.1409    -1.10    0.2721
## Year2005          -0.0607      0.1386    -0.44    0.6613
## Year2006          -0.1683      0.1350    -1.25    0.2129
## Year2007          -0.3434      0.1350    -2.54    0.0112 *
## Year2008          -0.3064      0.1281    -2.39    0.0170 *
## Year2009          -0.2424      0.1389    -1.75    0.0813 .
## Year2010          -0.1102      0.1370    -0.80    0.4213
## Year2011          -0.2676      0.1307    -2.05    0.0409 *
## Year2012          -0.2657      0.1319    -2.01    0.0443 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.215, Adjusted R-squared:  0.193
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 393 is an outlier with |weight| = 0 ( < 0.00012);
## 67 weights are ~= 1. The remaining 733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.230  0.865  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          1.25e-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 1.107 1          1.052
## LastAuthorFemale  1.063 1          1.031
## Year              1.174 16          1.005

```

Residuals from first and last author



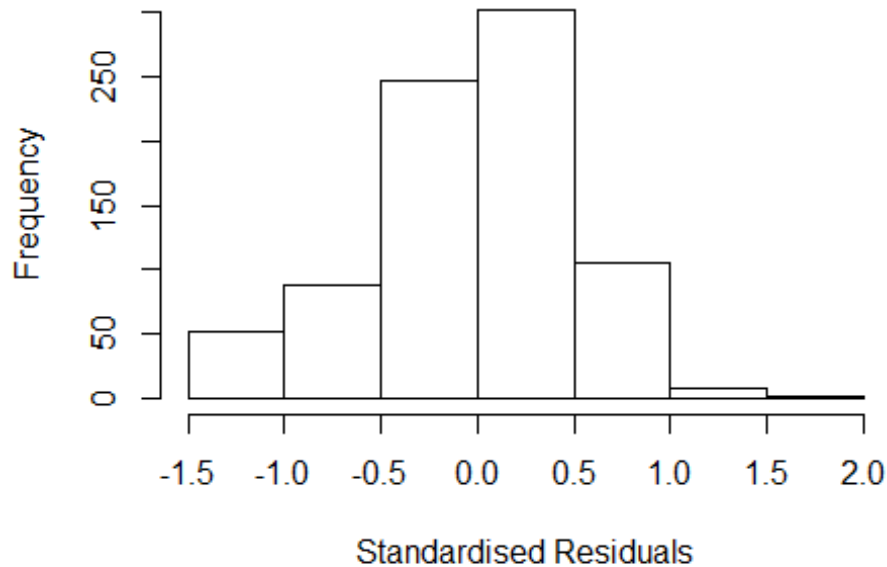
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5162 -0.3133 0.0268 0.3493 1.6047
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.39932 0.13956 10.03 <2e-16 ***
## FirstAuthorFemale1 0.11771 0.03853 3.05 0.0023 **
## LastAuthorFemale1 0.05367 0.04140 1.30 0.1952
## Year1997 0.08771 0.16378 0.54 0.5924
## Year1998 0.07809 0.16842 0.46 0.6430
## Year1999 -0.07639 0.17863 -0.43 0.6690
## Year2000 -0.14600 0.17885 -0.82 0.4146
## Year2001 -0.28472 0.20481 -1.39 0.1649
## Year2002 0.00592 0.16621 0.04 0.9716
## Year2003 -0.09721 0.16363 -0.59 0.5526
## Year2004 -0.20415 0.15720 -1.30 0.1944
## Year2005 -0.08446 0.16192 -0.52 0.6021
```

```

## Year2006      -0.12135      0.15372      -0.79      0.4301
## Year2007      -0.28301      0.15488      -1.83      0.0680 .
## Year2008      -0.26456      0.14760      -1.79      0.0735 .
## Year2009      -0.22530      0.15216      -1.48      0.1391
## Year2010      -0.05453      0.15244      -0.36      0.7207
## Year2011      -0.23229      0.15046      -1.54      0.1230
## Year2012      -0.23235      0.15229      -1.53      0.1275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0569, Adjusted R-squared:  0.0352
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 745 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.255  0.864  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      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 1.111 1      1.054
## Year      1.111 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
## -1.4986 -0.3270  0.0221  0.3483  1.5901
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4149    0.1402   10.09  <2e-16 ***
## FirstAuthorFemale1 0.1218    0.0388    3.14  0.0018 **
## Year1997          0.0837    0.1643    0.51  0.6106
## Year1998          0.0730    0.1691    0.43  0.6660
## Year1999         -0.0833    0.1808   -0.46  0.6452
## Year2000         -0.1481    0.1784   -0.83  0.4068
## Year2001         -0.2868    0.2091   -1.37  0.1705
## Year2002         -0.0021    0.1664   -0.01  0.9899
## Year2003         -0.0961    0.1627   -0.59  0.5550
## Year2004         -0.2120    0.1579   -1.34  0.1797
## Year2005         -0.0806    0.1625   -0.50  0.6198
## Year2006         -0.1228    0.1547   -0.79  0.4276
```

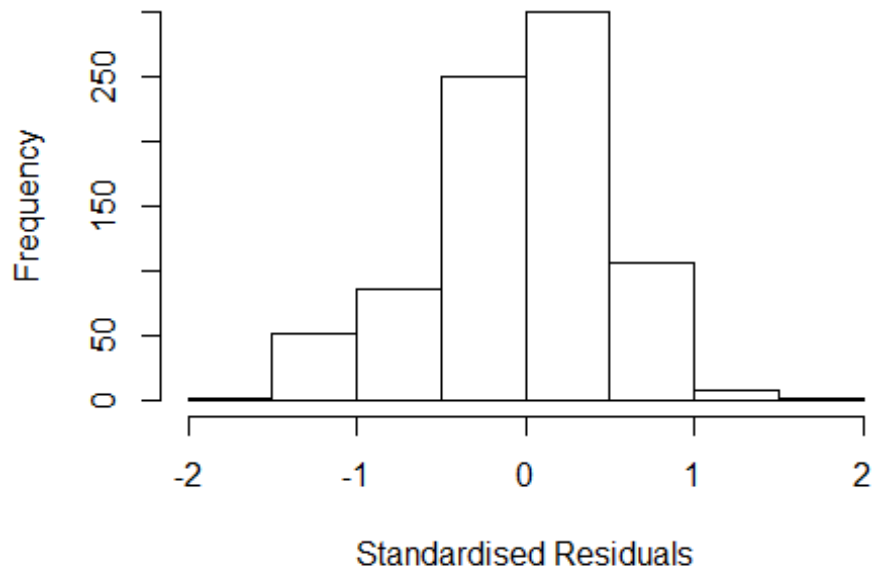


```

## Year2007          -0.2840      0.1558   -1.82   0.0687 .
## Year2008          -0.2602      0.1486   -1.75   0.0803 .
## Year2009          -0.2294      0.1532   -1.50   0.1346
## Year2010          -0.0580      0.1532   -0.38   0.7050
## Year2011          -0.2317      0.1511   -1.53   0.1256
## Year2012          -0.2359      0.1528   -1.54   0.1232
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0543, Adjusted R-squared:  0.0338
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.267  0.862  0.951   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      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 1.06 1          1.029
## Year            1.06 16          1.002

```

Residuals from last author



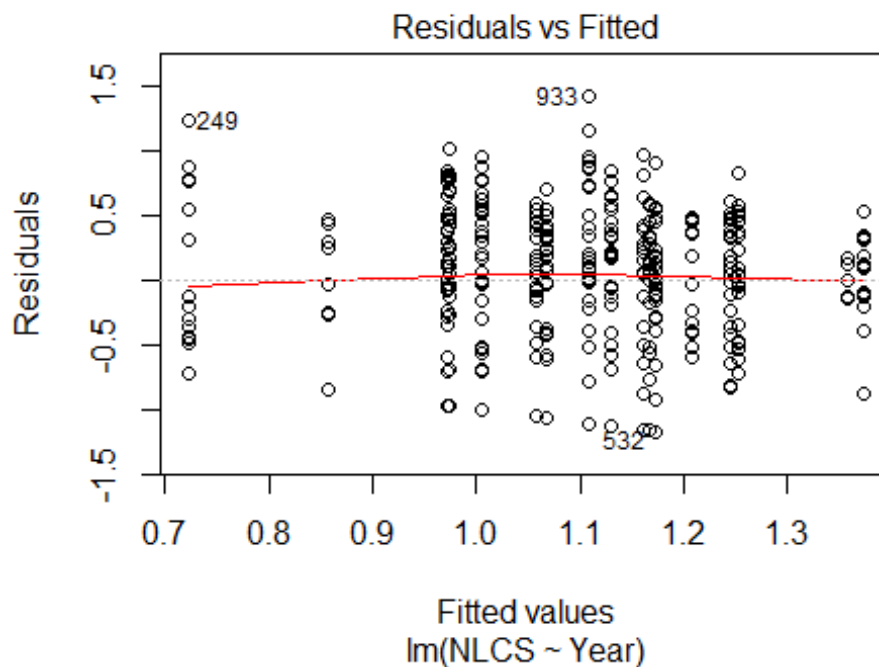
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5227 -0.3175 0.0232 0.3372 1.5733
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4303 0.1364 10.49 <2e-16 ***
## LastAuthorFemale1 0.0636 0.0416 1.53 0.126
## Year1997 0.0924 0.1607 0.57 0.566
## Year1998 0.0874 0.1710 0.51 0.609
## Year1999 -0.0827 0.1777 -0.47 0.642
## Year2000 -0.1584 0.1762 -0.90 0.369
## Year2001 -0.2902 0.2094 -1.39 0.166
## Year2002 0.0253 0.1627 0.16 0.877
## Year2003 -0.0798 0.1643 -0.49 0.628
## Year2004 -0.2059 0.1548 -1.33 0.184
## Year2005 -0.0534 0.1576 -0.34 0.735
## Year2006 -0.1027 0.1510 -0.68 0.497
```

```

## Year2007          -0.2826      0.1528    -1.85      0.065 .
## Year2008          -0.2645      0.1454    -1.82      0.069 .
## Year2009          -0.2158      0.1503    -1.44      0.151
## Year2010          -0.0305      0.1500    -0.20      0.839
## Year2011          -0.2151      0.1480    -1.45      0.146
## Year2012          -0.2110      0.1506    -1.40      0.162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.49
## Multiple R-squared:  0.0459, Adjusted R-squared:  0.0252
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 739 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.282  0.859  0.953  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      1.25e-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: 801"
## [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
##   66   56   42   40   99   28   45   26   34   36   37   27   33   44   46
## 2011 2012
##   53   47
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22    9   16   18   18    5   28   16   25   28   27   21   19   28   27
## 2011 2012

```

```
## 39 34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 8 15 16 15 5 26 15 24 23 24 21 15 24 25
## 2011 2012
## 35 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 = 35, df = 16, p-value = 0.004
```



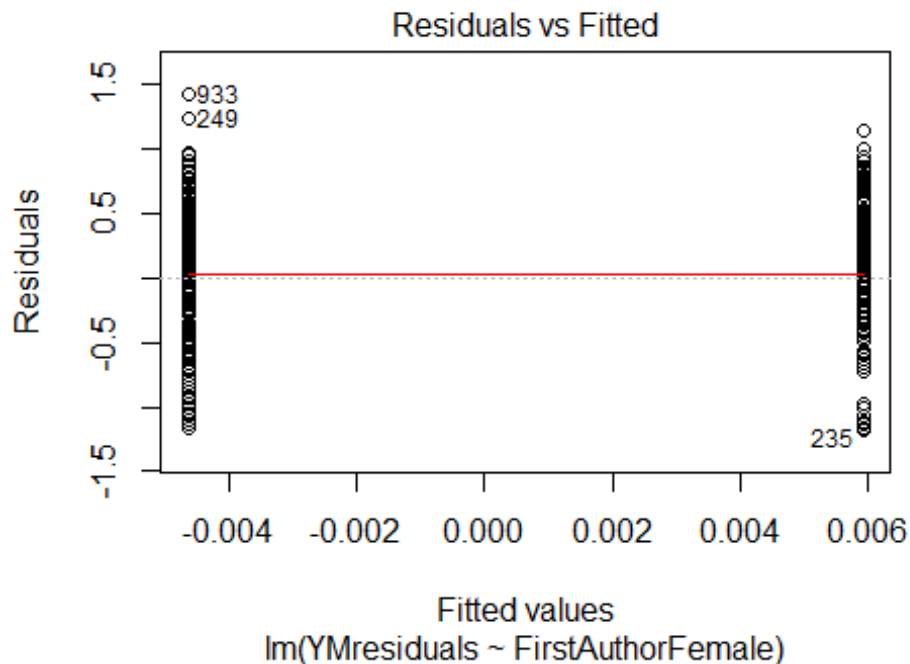
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.18, df = 1, p-value = 0.7

## [1] "Female first author team size 2018 geometric mean: 3.63424118566428"
## [1] "Male first author team size 2018 geometric mean: 5.15567787333116"

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

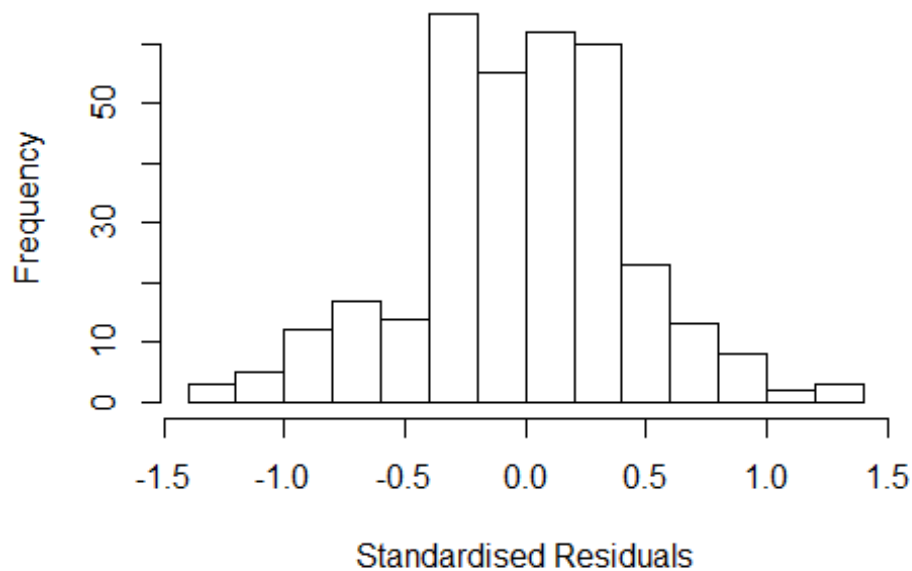
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.46169260907824"
## [1] "Male last author team size 2018 geometric mean: 4.51027919721931"

## 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 = 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 1.653 1      1.286
## LastAuthorFemale  1.974 1      1.405
## UniqueAuthors    6.810 4      1.271
## Year             10.829 16     1.077
```

Residuals from first and last author and team size



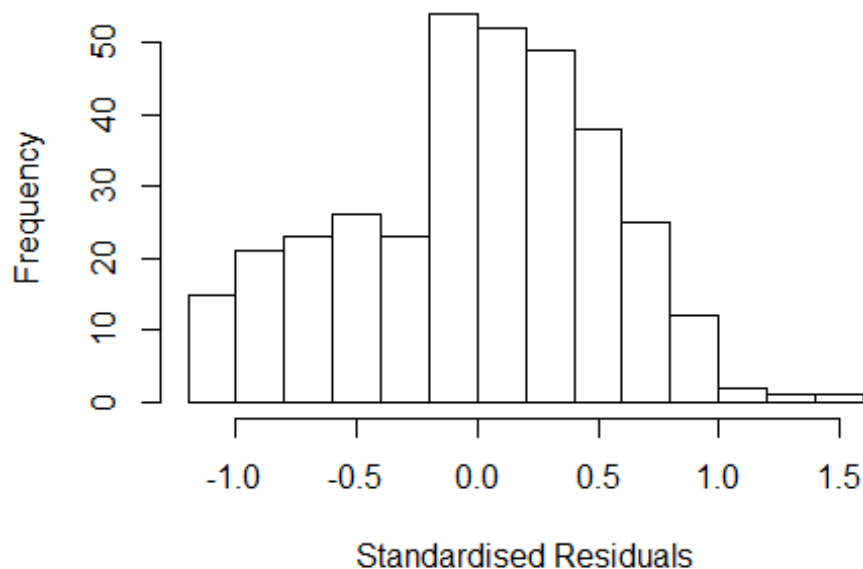
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.310554 -0.258421 0.000698 0.281357 1.224575
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.46780 0.14320 3.27 0.0012 **
## FirstAuthorFemale1 -0.00574 0.05897 -0.10 0.9226
## LastAuthorFemale1 -0.04765 0.07603 -0.63 0.5313
## UniqueAuthors2 0.55966 0.16933 3.31 0.0011 **
## UniqueAuthors3 0.79610 0.11143 7.14 6.2e-12 ***
## UniqueAuthors4 0.71731 0.11949 6.00 5.3e-09 ***
## UniqueAuthors5 1.04760 0.09281 11.29 < 2e-16 ***
## Year1997 -0.41549 0.12814 -3.24 0.0013 **
## Year1998 0.08226 0.14245 0.58 0.5641
## Year1999 -0.02100 0.19222 -0.11 0.9131
```

```

## Year2000      -0.37736    0.20424   -1.85    0.0656 .
## Year2001      0.16628    0.13401    1.24    0.2156
## Year2002     -0.17820    0.12687   -1.40    0.1611
## Year2003     -0.11799    0.15519   -0.76    0.4476
## Year2004     -0.09172    0.14622   -0.63    0.5309
## Year2005     -0.17306    0.14465   -1.20    0.2324
## Year2006     -0.28513    0.15029   -1.90    0.0587 .
## Year2007      0.00109    0.15764    0.01    0.9945
## Year2008     -0.23358    0.18995   -1.23    0.2197
## Year2009     -0.18713    0.18746   -1.00    0.3189
## Year2010     -0.20485    0.14277   -1.43    0.1523
## Year2011     -0.11820    0.13990   -0.84    0.3988
## Year2012     -0.19957    0.15037   -1.33    0.1854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.433, Adjusted R-squared:  0.394
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.250  0.857  0.951  0.884  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      2.92e-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 1.199 1      1.095
## LastAuthorFemale  1.346 1      1.160
## Year              1.558 16      1.014

```

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
## -1.1887 -0.3854 0.0431 0.3606 1.4654
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27959 0.11635 11.00 <2e-16 ***
## FirstAuthorFemale1 0.00866 0.06391 0.14 0.892
## LastAuthorFemale1 -0.15811 0.10390 -1.52 0.129
## Year1997 -0.40768 0.18899 -2.16 0.032 *
## Year1998 0.15495 0.13997 1.11 0.269
## Year1999 -0.09959 0.16433 -0.61 0.545
## Year2000 -0.57741 0.23596 -2.45 0.015 *
## Year2001 0.10697 0.12830 0.83 0.405
## Year2002 -0.12137 0.14116 -0.86 0.391
## Year2003 -0.00869 0.15964 -0.05 0.957
## Year2004 -0.15014 0.14973 -1.00 0.317
## Year2005 -0.02910 0.14184 -0.21 0.838
```

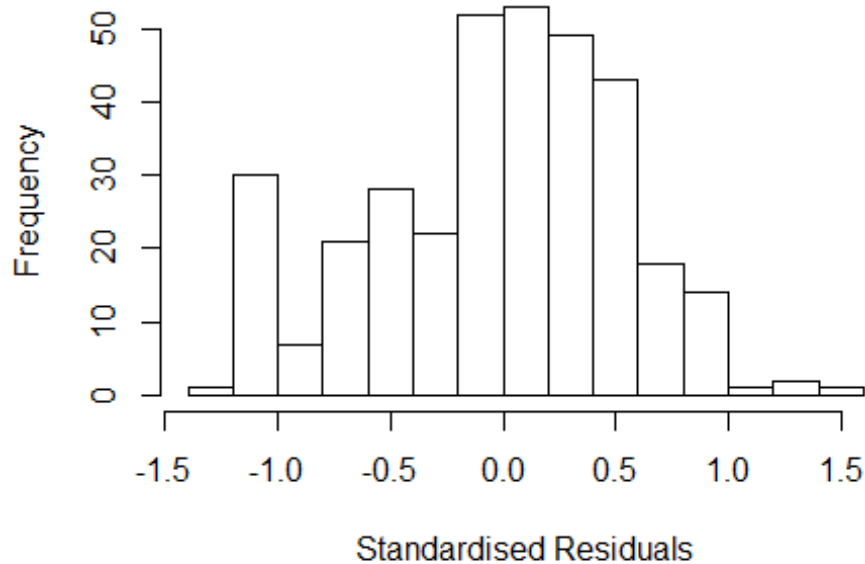


```

## Year2006      -0.34877    0.16518   -2.11    0.036 *
## Year2007      0.01961    0.15342    0.13    0.898
## Year2008     -0.20108    0.22321   -0.90    0.368
## Year2009     -0.21997    0.23071   -0.95    0.341
## Year2010     -0.13694    0.16345   -0.84    0.403
## Year2011     -0.19055    0.17634   -1.08    0.281
## Year2012     -0.21599    0.19978   -1.08    0.280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0883, Adjusted R-squared:  0.0375
## 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.
##  0.373  0.835   0.940   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      2.92e-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 1.188 1      1.090
## Year      1.188 16      1.005

```

Residuals from first author



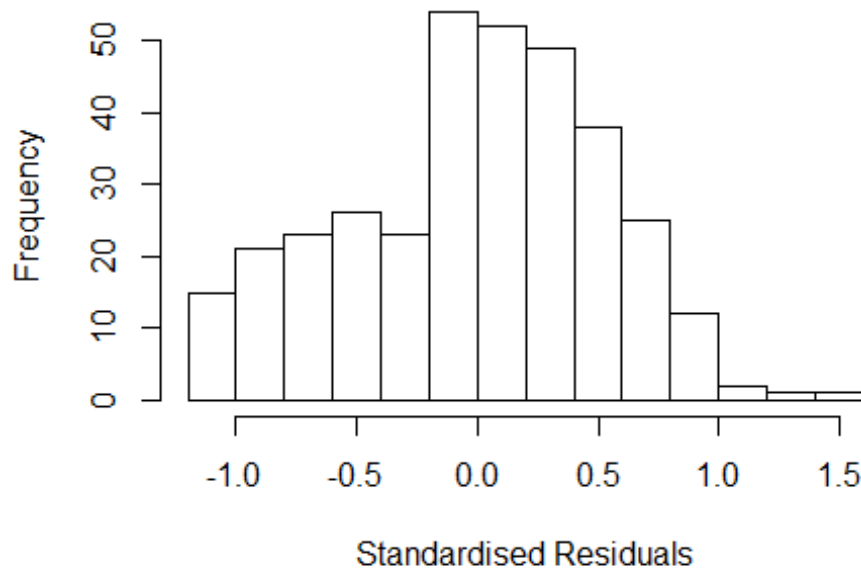
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2130 -0.4071 0.0523 0.3659 1.5268
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25067 0.11235 11.13 <2e-16 ***
## FirstAuthorFemale1 0.00132 0.06450 0.02 0.984
## Year1997 -0.41634 0.18273 -2.28 0.023 *
## Year1998 0.16387 0.13668 1.20 0.231
## Year1999 -0.08954 0.15778 -0.57 0.571
## Year2000 -0.59623 0.24966 -2.39 0.018 *
## Year2001 0.10674 0.12203 0.87 0.382
## Year2002 -0.12380 0.13718 -0.90 0.367
## Year2003 0.00307 0.15699 0.02 0.984
## Year2004 -0.12720 0.14400 -0.88 0.378
## Year2005 -0.03765 0.13907 -0.27 0.787
## Year2006 -0.33158 0.16022 -2.07 0.039 *
```

```

## Year2007          0.01027    0.15096    0.07    0.946
## Year2008         -0.17724    0.22389   -0.79    0.429
## Year2009         -0.22115    0.23787   -0.93    0.353
## Year2010         -0.12703    0.16186   -0.78    0.433
## Year2011         -0.19404    0.17173   -1.13    0.259
## Year2012         -0.24844    0.19540   -1.27    0.204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.0762, Adjusted R-squared:  0.0277
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.342  0.853   0.938   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      2.92e-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 1.312 1          1.145
## Year            1.312 16          1.009

```

Residuals from last author



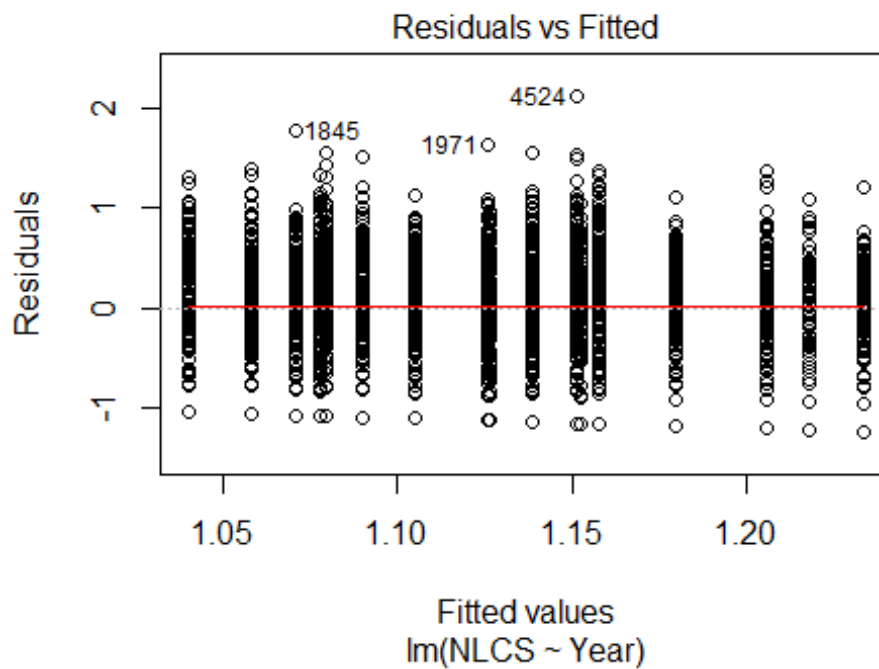
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1823 -0.3824 0.0435 0.3636 1.4616
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28292 0.11123 11.53 <2e-16 ***
## LastAuthorFemale1 -0.15815 0.10361 -1.53 0.128
## Year1997 -0.40772 0.18875 -2.16 0.031 *
## Year1998 0.15374 0.13927 1.10 0.270
## Year1999 -0.10061 0.16433 -0.61 0.541
## Year2000 -0.57520 0.23319 -2.47 0.014 *
## Year2001 0.10708 0.12780 0.84 0.403
## Year2002 -0.12164 0.14034 -0.87 0.387
## Year2003 -0.00705 0.15926 -0.04 0.965
## Year2004 -0.14889 0.14833 -1.00 0.316
## Year2005 -0.02956 0.14121 -0.21 0.834
## Year2006 -0.34805 0.16516 -2.11 0.036 *
```

```

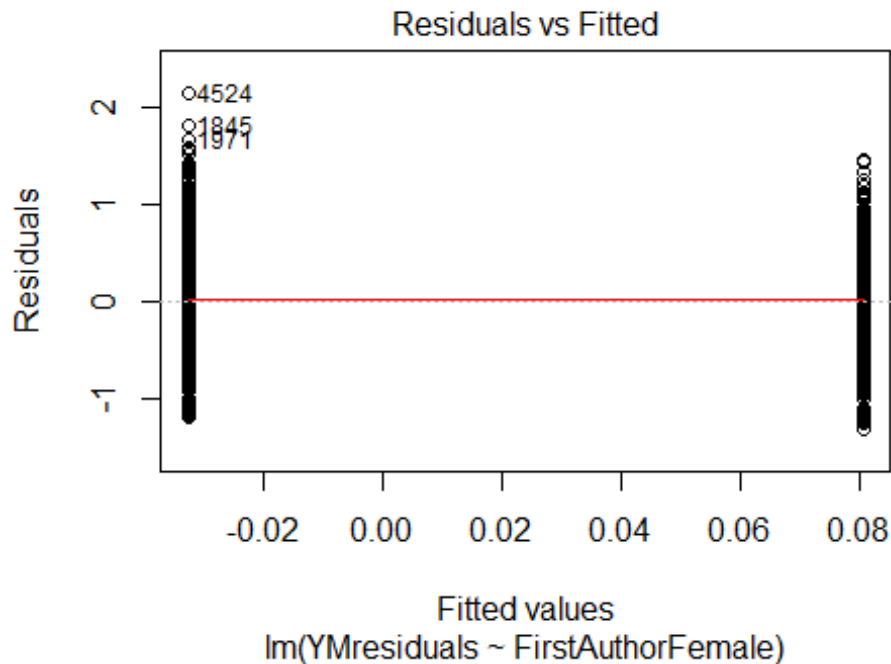
## Year2007      0.02065      0.15295      0.14      0.893
## Year2008     -0.19959      0.22175     -0.90      0.369
## Year2009     -0.22120      0.22880     -0.97      0.334
## Year2010     -0.13782      0.16281     -0.85      0.398
## Year2011     -0.19118      0.17533     -1.09      0.276
## Year2012     -0.21552      0.19852     -1.09      0.278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.088, Adjusted R-squared:  0.0401
## 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.
##  0.385  0.839  0.941  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      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 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
##  186  208  186  176  191  249  203  217  218  241  274  292  330  305  365
## 2011 2012
##  373  339
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   75  112   92   91  111  164  145  156  169  195  228  212  264  228  275
## 2011 2012

```

```
## 300 251
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 101 86 86 105 145 134 147 155 183 208 194 233 203 243
## 2011 2012
## 286 229
## [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
```

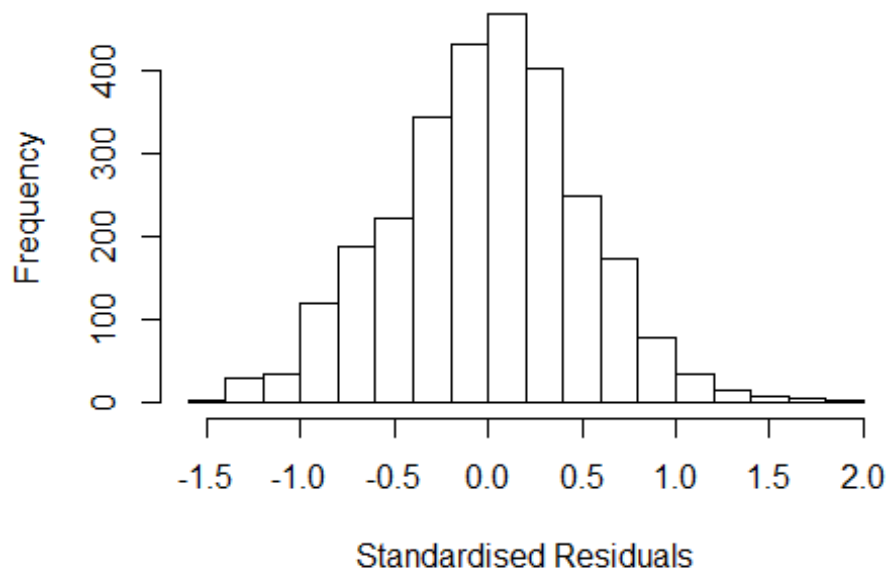


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 25, df = 1, p-value = 6e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 5.70444882750511"
## [1] "Male first author team size 2018 geometric mean: 5.01309685537012"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.08253636076448"
## [1] "Male last author team size 2018 geometric mean: 5.26096588045492"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6800, 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 1.061 1          1.030
## LastAuthorFemale  1.048 1          1.024
## UniqueAuthors    1.223 4          1.026
## Year             1.220 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
## -1.4767 -0.3254 0.0184 0.3290 1.9650
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.81417 0.07460 10.91 < 2e-16 ***
## FirstAuthorFemale1 0.07938 0.02100 3.78 0.00016 ***
## LastAuthorFemale1 0.12532 0.02708 4.63 3.9e-06 ***
## UniqueAuthors2 0.18870 0.05387 3.50 0.00047 ***
## UniqueAuthors3 0.29398 0.04890 6.01 2.1e-09 ***
## UniqueAuthors4 0.37533 0.04717 7.96 2.5e-15 ***
## UniqueAuthors5 0.59903 0.04371 13.71 < 2e-16 ***
## Year1997 0.00334 0.08274 0.04 0.96782
## Year1998 -0.05586 0.08340 -0.67 0.50306
## Year1999 -0.01757 0.07962 -0.22 0.82538
```

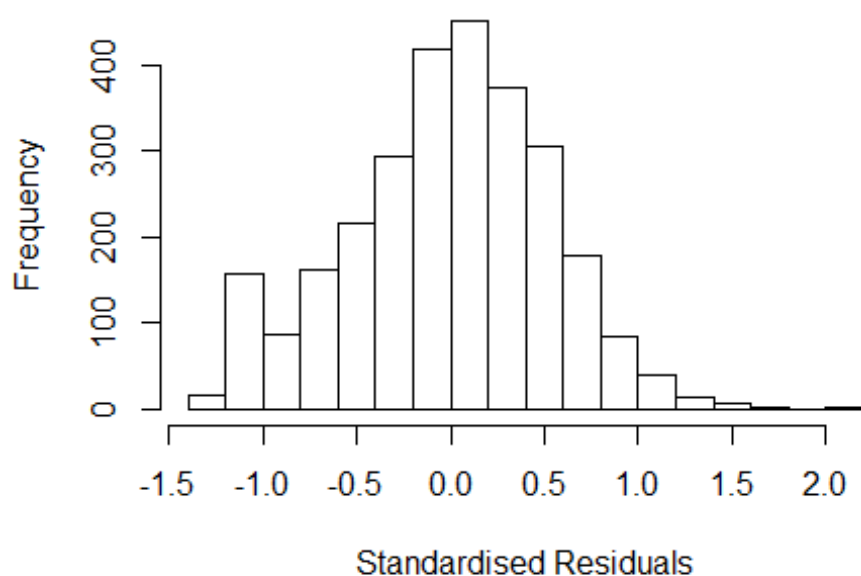


```

## Year2000      -0.07651    0.08004   -0.96  0.33920
## Year2001      -0.19966    0.07652   -2.61  0.00913 **
## Year2002      -0.11225    0.07753   -1.45  0.14777
## Year2003      -0.21702    0.07645   -2.84  0.00456 **
## Year2004      -0.13243    0.07592   -1.74  0.08119 .
## Year2005      -0.16995    0.07211   -2.36  0.01850 *
## Year2006      -0.17879    0.07266   -2.46  0.01393 *
## Year2007      -0.14117    0.07337   -1.92  0.05443 .
## Year2008      -0.20760    0.07176   -2.89  0.00384 **
## Year2009      -0.19187    0.07772   -2.47  0.01361 *
## Year2010      -0.20948    0.07081   -2.96  0.00312 **
## Year2011      -0.11624    0.07117   -1.63  0.10253
## Year2012      -0.16150    0.07392   -2.18  0.02899 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.151, Adjusted R-squared:  0.144
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 237 weights are ~= 1. The remaining 2568 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0725 0.8640 0.9510 0.9030 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 1.048 1 1.024
## LastAuthorFemale 1.037 1 1.018
## Year 1.042 16 1.001

```

Residuals from first and last author



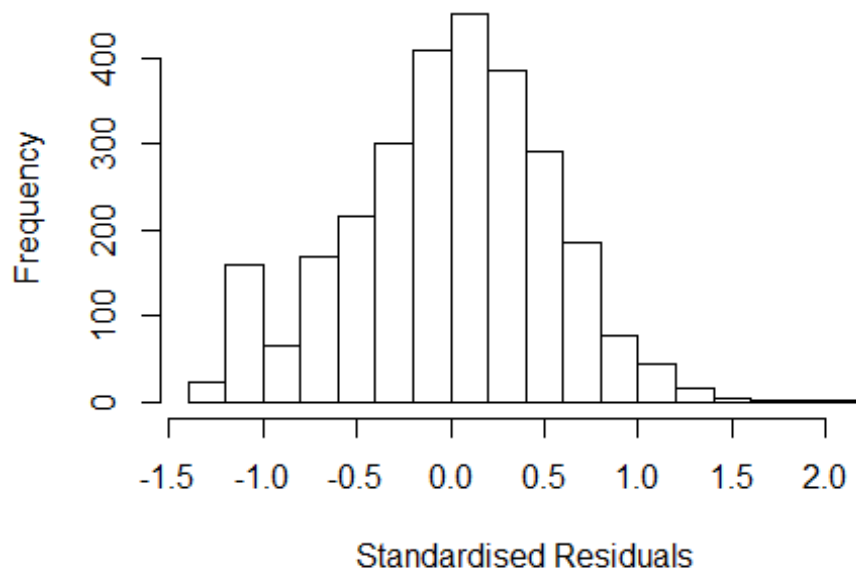
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3581 -0.3437 0.0272 0.3520 2.1263
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21245 0.05908 20.52 < 2e-16 ***
## FirstAuthorFemale1 0.11112 0.02199 5.05 4.6e-07 ***
## LastAuthorFemale1 0.09588 0.02860 3.35 0.00081 ***
## Year1997 -0.01280 0.08060 -0.16 0.87379
## Year1998 -0.10530 0.08195 -1.28 0.19891
## Year1999 0.00198 0.07938 0.02 0.98011
## Year2000 -0.10442 0.07638 -1.37 0.17172
## Year2001 -0.21257 0.07573 -2.81 0.00504 **
## Year2002 -0.06138 0.07299 -0.84 0.40048
## Year2003 -0.20576 0.07553 -2.72 0.00649 **
## Year2004 -0.12785 0.07387 -1.73 0.08362 .
## Year2005 -0.13406 0.07088 -1.89 0.05868 .
```

```

## Year2006      -0.17079    0.07035   -2.43  0.01525 *
## Year2007      -0.16150    0.07026   -2.30  0.02159 *
## Year2008      -0.19778    0.06919   -2.86  0.00429 **
## Year2009      -0.16378    0.07593   -2.16  0.03109 *
## Year2010      -0.15990    0.06763   -2.36  0.01814 *
## Year2011      -0.07678    0.06860   -1.12  0.26312
## Year2012      -0.10242    0.07119   -1.44  0.15035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0246, Adjusted R-squared:  0.0183
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 224 weights are ~= 1. The remaining 2581 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0578 0.8590 0.9490 0.9020 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      3.57e-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 1.027 1      1.014
## Year      1.027 16      1.001

```

Residuals from first author



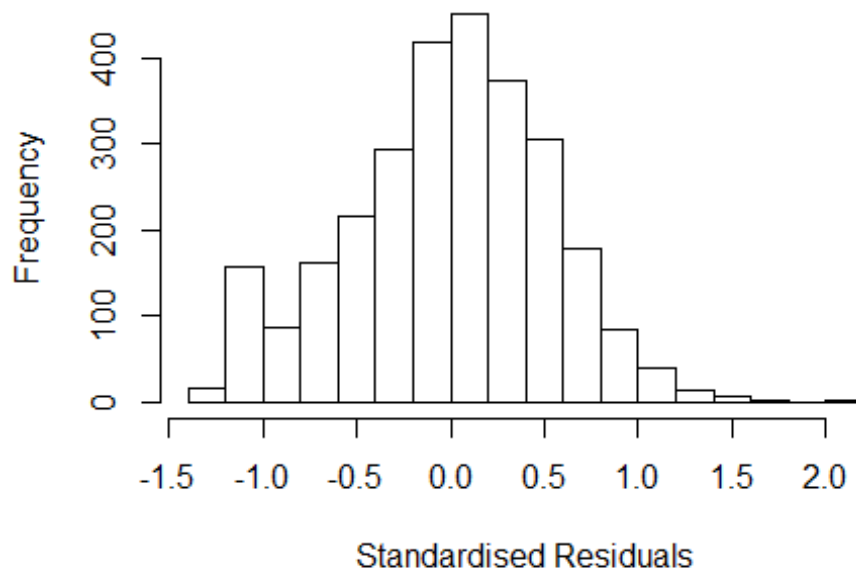
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3504 -0.3436 0.0286 0.3460 2.1171
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21685 0.05866 20.74 < 2e-16 ***
## FirstAuthorFemale1 0.12343 0.02173 5.68 1.5e-08 ***
## Year1997 -0.00535 0.08048 -0.07 0.9470
## Year1998 -0.09958 0.08181 -1.22 0.2236
## Year1999 0.01010 0.07897 0.13 0.8982
## Year2000 -0.09602 0.07583 -1.27 0.2055
## Year2001 -0.20584 0.07540 -2.73 0.0064 **
## Year2002 -0.05816 0.07314 -0.80 0.4266
## Year2003 -0.19846 0.07533 -2.63 0.0085 **
## Year2004 -0.11775 0.07356 -1.60 0.1095
## Year2005 -0.13124 0.07054 -1.86 0.0629 .
## Year2006 -0.16622 0.07021 -2.37 0.0180 *
```

```

## Year2007          -0.15399    0.06990   -2.20    0.0277 *
## Year2008          -0.19232    0.06879   -2.80    0.0052 **
## Year2009          -0.15752    0.07543   -2.09    0.0369 *
## Year2010          -0.15241    0.06732   -2.26    0.0236 *
## Year2011          -0.07196    0.06837   -1.05    0.2927
## Year2012          -0.09487    0.07081   -1.34    0.1804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0208, Adjusted R-squared:  0.0149
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 226 weights are ~= 1. The remaining 2579 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0616 0.8620 0.9500 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      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 1.013 1          1.007
## Year              1.013 16          1.000

```

Residuals from last author



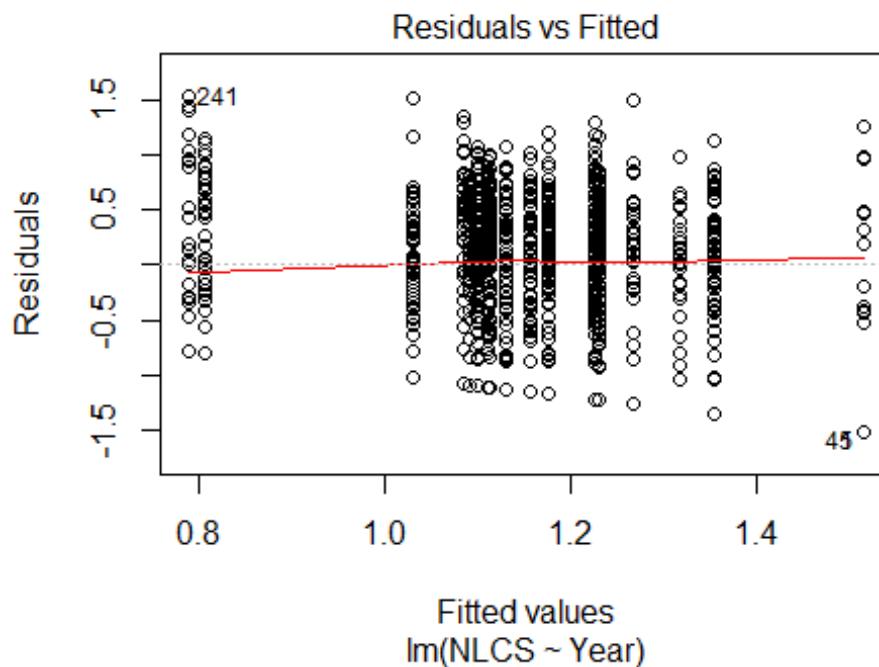
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2987 -0.3515  0.0328  0.3622  2.1008
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23009    0.05951   20.67  < 2e-16 ***
## LastAuthorFemale1 0.11950    0.02815    4.24  2.3e-05 ***
## Year1997       -0.00922    0.08086   -0.11  0.9092
## Year1998       -0.09824    0.08243   -1.19  0.2334
## Year1999        0.00652    0.07991    0.08  0.9350
## Year2000       -0.08630    0.07679   -1.12  0.2612
## Year2001       -0.19986    0.07631   -2.62  0.0089 **
## Year2002       -0.05093    0.07315   -0.70  0.4863
## Year2003       -0.18840    0.07634   -2.47  0.0136 *
## Year2004       -0.12069    0.07462   -1.62  0.1059
## Year2005       -0.11750    0.07144   -1.64  0.1001
## Year2006       -0.16064    0.07090   -2.27  0.0235 *
```

```

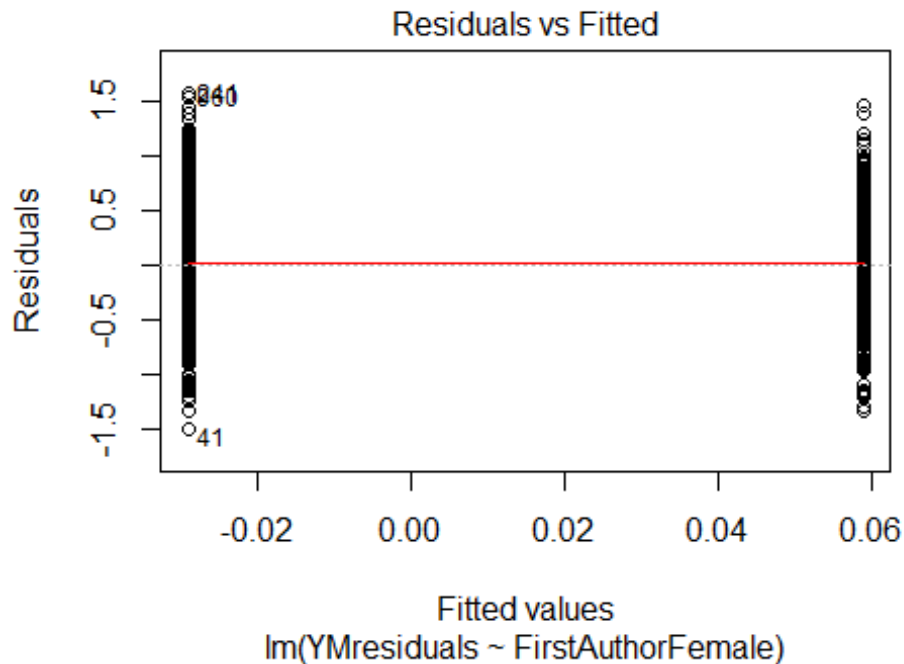
## Year2007          -0.14969      0.07112      -2.10      0.0354 *
## Year2008          -0.17963      0.06947      -2.59      0.0098 **
## Year2009          -0.14858      0.07648      -1.94      0.0521 .
## Year2010          -0.14313      0.06813      -2.10      0.0357 *
## Year2011          -0.06888      0.06912      -1.00      0.3190
## Year2012          -0.08973      0.07162      -1.25      0.2104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.016, Adjusted R-squared:  0.01
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 225 weights are ~= 1. The remaining 2580 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0745 0.8670 0.9500 0.9030 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: 2805"
## [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
##   68   77   80   99  109  125  121  115   62   95  116  117  139  152  168
## 2011 2012
##  194  166
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   27   37   55   61   73   82   79   36   59   71   75   98   86   88
## 2011 2012

```

```
## 126 104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 23 32 52 52 67 74 73 31 50 57 72 85 81 81
## 2011 2012
## 114 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 = 34, df = 16, p-value = 0.006
```



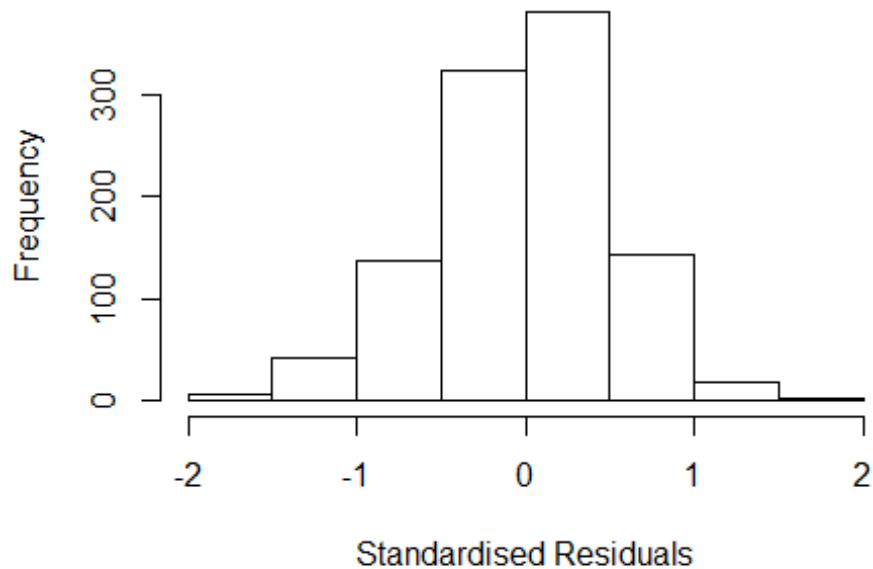
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.7, df = 1, p-value = 0.02
## [1] "Female first author team size 2018 geometric mean: 4.34371377295208"
## [1] "Male first author team size 2018 geometric mean: 4.21617056072211"
## 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 = 1100, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.15190227854665"
## [1] "Male last author team size 2018 geometric mean: 4.3394974418459"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 930, 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	1.163	1	1.078
LastAuthorFemale	1.206	1	1.098
UniqueAuthors	1.936	4	1.086
Year	2.052	16	1.023

Residuals from first and last author and team size



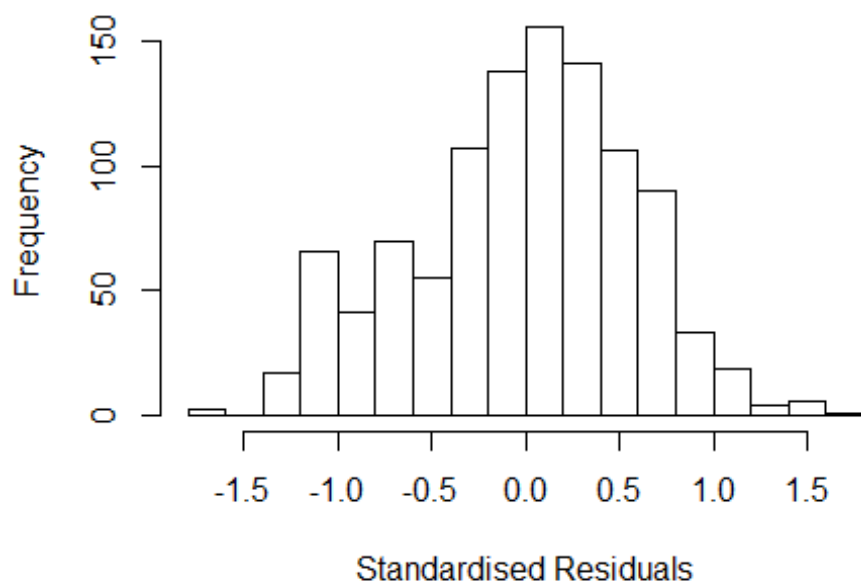
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5691 -0.3508 0.0203 0.3570 1.9453
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0596 0.2363 4.48 8.1e-06 ***
## FirstAuthorFemale1 0.0432 0.0368 1.17 0.2406
## LastAuthorFemale1 0.1038 0.0389 2.66 0.0078 **
## UniqueAuthors2 0.3694 0.0719 5.14 3.4e-07 ***
## UniqueAuthors3 0.5378 0.0687 7.83 1.2e-14 ***
## UniqueAuthors4 0.6143 0.0754 8.15 1.1e-15 ***
## UniqueAuthors5 0.9295 0.0619 15.01 < 2e-16 ***
## Year1997 -0.3963 0.2458 -1.61 0.1072
## Year1998 -0.6495 0.2629 -2.47 0.0137 *
## Year1999 -0.4675 0.2547 -1.84 0.0667 .
```

```

## Year2000          -0.7055      0.2385    -2.96    0.0032 **
## Year2001          -0.6119      0.2331    -2.62    0.0088 **
## Year2002          -0.5631      0.2334    -2.41    0.0160 *
## Year2003          -0.4834      0.2323    -2.08    0.0377 *
## Year2004          -0.4253      0.2596    -1.64    0.1016
## Year2005          -0.6413      0.2410    -2.66    0.0079 **
## Year2006          -0.5759      0.2359    -2.44    0.0148 *
## Year2007          -0.5850      0.2324    -2.52    0.0120 *
## Year2008          -0.5253      0.2315    -2.27    0.0235 *
## Year2009          -0.4928      0.2317    -2.13    0.0336 *
## Year2010          -0.6101      0.2330    -2.62    0.0090 **
## Year2011          -0.5586      0.2304    -2.42    0.0155 *
## Year2012          -0.5670      0.2323    -2.44    0.0148 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.298, Adjusted R-squared:  0.283
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 969 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0905 0.8710 0.9470 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          9.51e-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 1.131 1          1.064
## LastAuthorFemale 1.136 1          1.066
## Year              1.214 16          1.006

```

Residuals from first and last author



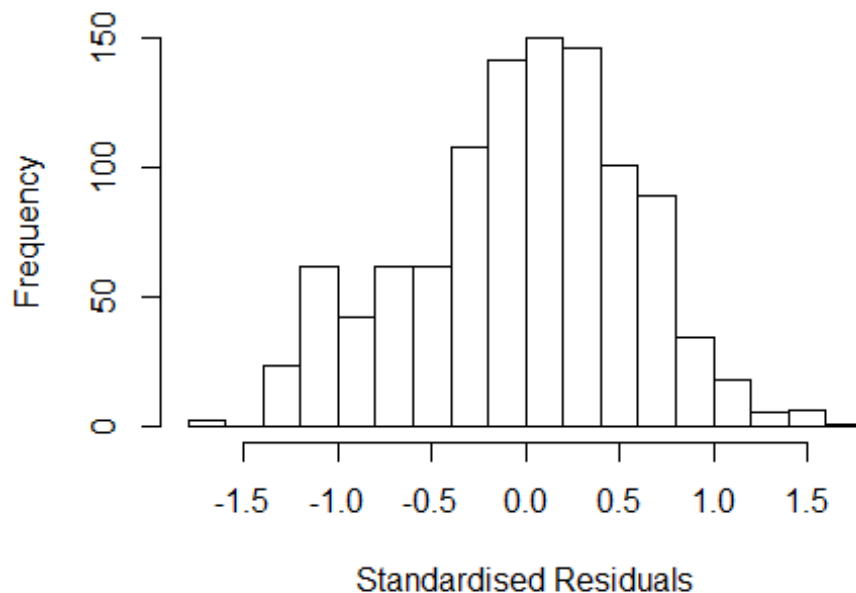
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6914 -0.3789  0.0278  0.3877  1.6420
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6914     0.2596   6.52 1.1e-10 ***
## FirstAuthorFemale1  0.0653     0.0401   1.63  0.10393
## LastAuthorFemale1  0.0728     0.0408   1.78  0.07469 .
## Year1997          -0.3779     0.2816  -1.34  0.17986
## Year1998          -1.0124     0.2968  -3.41  0.00067 ***
## Year1999          -0.6691     0.2947  -2.27  0.02338 *
## Year2000          -0.9914     0.2823  -3.51  0.00046 ***
## Year2001          -0.7398     0.2656  -2.79  0.00545 **
## Year2002          -0.6334     0.2683  -2.36  0.01842 *
## Year2003          -0.6371     0.2673  -2.38  0.01735 *
## Year2004          -0.4072     0.2848  -1.43  0.15298
## Year2005          -0.6295     0.2750  -2.29  0.02229 *
```

```

## Year2006          -0.5579      0.2724    -2.05  0.04082 *
## Year2007          -0.5397      0.2683    -2.01  0.04450 *
## Year2008          -0.4425      0.2671    -1.66  0.09791 .
## Year2009          -0.3854      0.2655    -1.45  0.14689
## Year2010          -0.5767      0.2693    -2.14  0.03248 *
## Year2011          -0.5297      0.2650    -2.00  0.04587 *
## Year2012          -0.4969      0.2662    -1.87  0.06221 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0817, Adjusted R-squared:  0.0657
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 955 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.328  0.855   0.948   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      9.51e-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 1.106 1          1.052
## Year              1.106 16          1.003

```

Residuals from first author



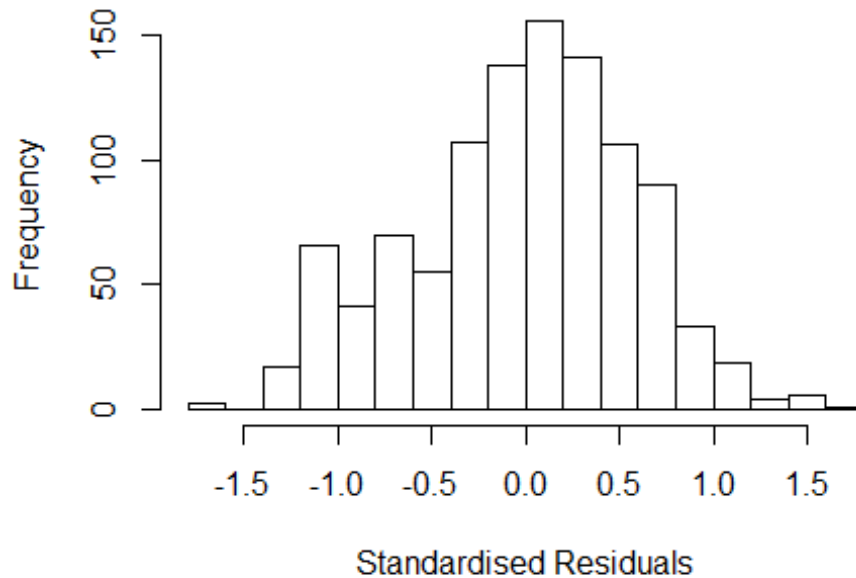
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6894 -0.3789 0.0278 0.3857 1.6294
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6894 0.2635 6.41 2.2e-10 ***
## FirstAuthorFemale1 0.0780 0.0396 1.97 0.04924 *
## Year1997 -0.3630 0.2853 -1.27 0.20353
## Year1998 -0.9978 0.2995 -3.33 0.00089 ***
## Year1999 -0.6556 0.2979 -2.20 0.02801 *
## Year2000 -0.9799 0.2864 -3.42 0.00065 ***
## Year2001 -0.7250 0.2693 -2.69 0.00722 **
## Year2002 -0.6245 0.2717 -2.30 0.02176 *
## Year2003 -0.6208 0.2708 -2.29 0.02211 *
## Year2004 -0.3822 0.2871 -1.33 0.18338
## Year2005 -0.6075 0.2781 -2.18 0.02915 *
## Year2006 -0.5299 0.2749 -1.93 0.05424 .
```

```

## Year2007          -0.5272      0.2719   -1.94  0.05283 .
## Year2008          -0.4236      0.2703   -1.57  0.11734
## Year2009          -0.3688      0.2691   -1.37  0.17084
## Year2010          -0.5619      0.2726   -2.06  0.03948 *
## Year2011          -0.5049      0.2680   -1.88  0.05985 .
## Year2012          -0.4816      0.2697   -1.79  0.07442 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0792, Adjusted R-squared:  0.0641
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 956 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.329  0.855  0.946  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      9.51e-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 1.108 1          1.053
## Year            1.108 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
## -1.7152 -0.3779 0.0421 0.3925 1.6334
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7152 0.2612 6.57 8.1e-11 ***
## LastAuthorFemale1 0.0860 0.0403 2.13 0.03307 *
## Year1997 -0.3917 0.2829 -1.38 0.16646
## Year1998 -1.0276 0.2989 -3.44 0.00061 ***
## Year1999 -0.6796 0.2959 -2.30 0.02185 *
## Year2000 -1.0057 0.2850 -3.53 0.00044 ***
## Year2001 -0.7463 0.2676 -2.79 0.00538 **
## Year2002 -0.6350 0.2703 -2.35 0.01899 *
## Year2003 -0.6428 0.2692 -2.39 0.01711 *
## Year2004 -0.4097 0.2861 -1.43 0.15241
## Year2005 -0.6320 0.2767 -2.28 0.02255 *
## Year2006 -0.5600 0.2742 -2.04 0.04139 *
```

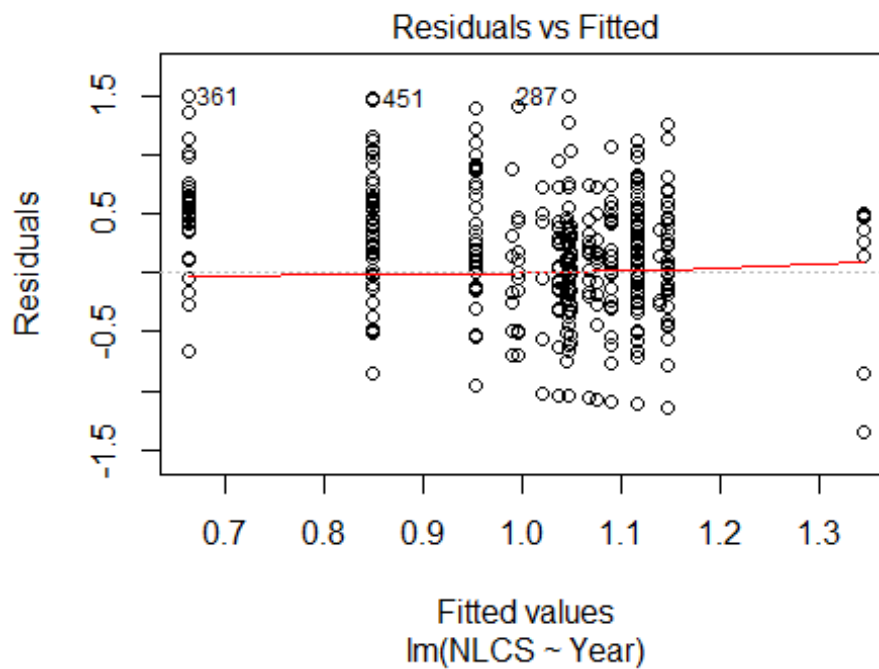


```

## Year2007          -0.5428      0.2704    -2.01  0.04494 *
## Year2008          -0.4529      0.2692    -1.68  0.09279 .
## Year2009          -0.3865      0.2674    -1.45  0.14868
## Year2010          -0.5773      0.2714    -2.13  0.03366 *
## Year2011          -0.5330      0.2671    -2.00  0.04628 *
## Year2012          -0.4963      0.2682    -1.85  0.06455 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.08,   Adjusted R-squared:  0.0648
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 964 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.314  0.854  0.947  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      9.51e-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: 1052"
## [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
##    7   11   13   17   22   21   24   14   19   35   31   55   79   95   58
## 2011 2012
##   91   87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    8    6   12   11   14   14    8   13   28   24   34   65   84   41
## 2011 2012

```

```
## 79 59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 7 6 12 10 12 13 5 13 28 23 31 64 79 40
## 2011 2012
## 69 52
## [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
```



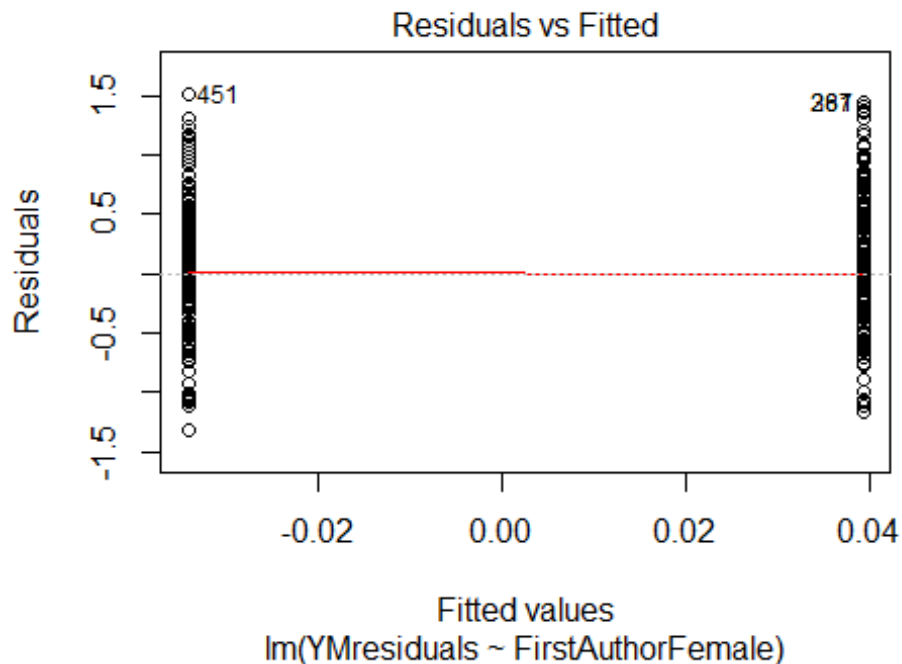
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.8, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 3.48780160757159"
## [1] "Male first author team size 2018 geometric mean: 3.00939025143113"

## 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 = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.55247420668859"
## [1] "Male last author team size 2018 geometric mean: 2.99165241522727"

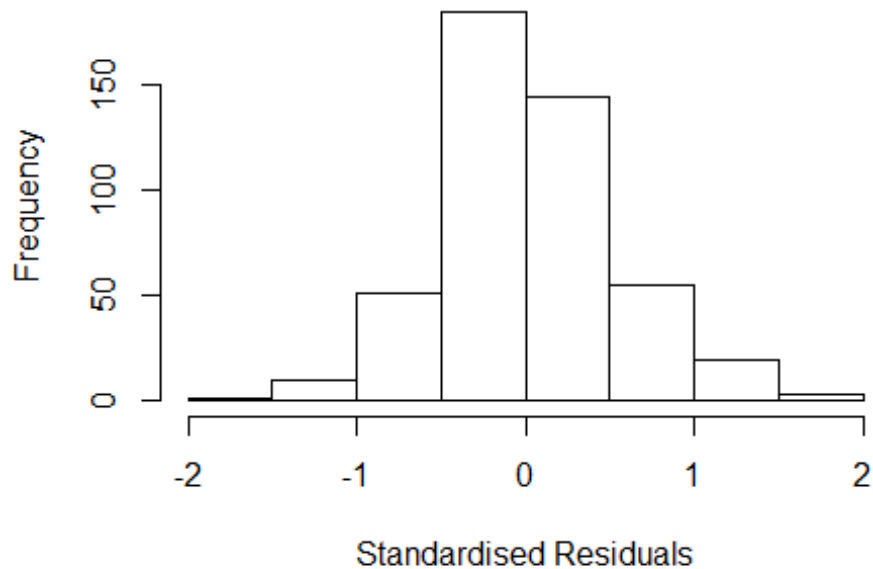
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 530, 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	1.633	1	1.278
LastAuthorFemale	1.598	1	1.264
UniqueAuthors	3.286	4	1.160
Year	5.731	16	1.056

Residuals from first and last author and team size



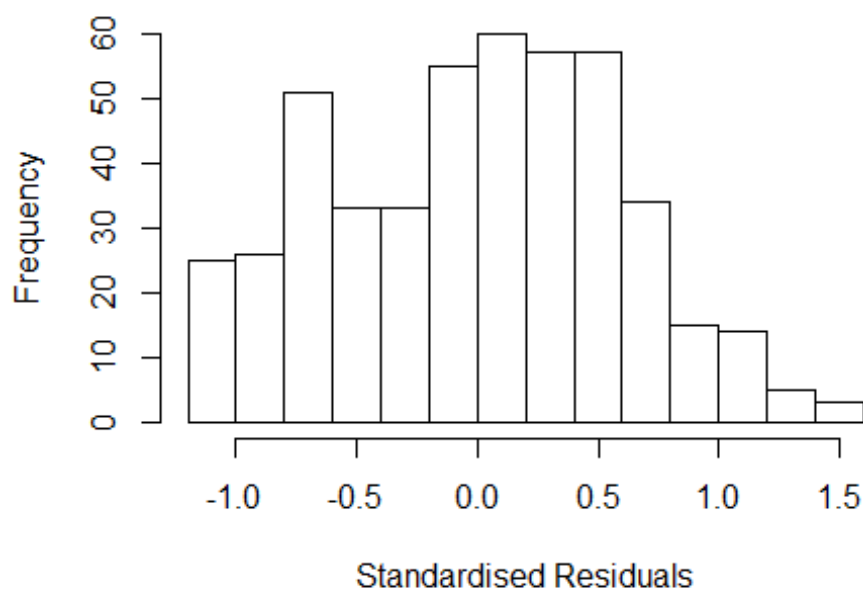
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7624 -0.3176 -0.0273 0.3217 1.5283
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3854 0.1509 2.55 0.011 *
## FirstAuthorFemale1 0.1323 0.0578 2.29 0.023 *
## LastAuthorFemale1 -0.1247 0.0572 -2.18 0.030 *
## UniqueAuthors2 0.6207 0.0945 6.57 1.4e-10 ***
## UniqueAuthors3 0.6478 0.0864 7.49 3.6e-13 ***
## UniqueAuthors4 0.7230 0.0913 7.92 1.9e-14 ***
## UniqueAuthors5 0.7911 0.0767 10.31 < 2e-16 ***
## Year1997 0.0216 0.1953 0.11 0.912
## Year1998 0.2955 0.2522 1.17 0.242
## Year1999 0.0487 0.1623 0.30 0.765
```

```

## Year2000      0.1694      0.1773      0.96      0.340
## Year2001      0.2303      0.2417      0.95      0.341
## Year2002      0.1004      0.1631      0.62      0.539
## Year2003      0.7564      0.7768      0.97      0.331
## Year2004      0.0523      0.1528      0.34      0.732
## Year2005      0.2353      0.1604      1.47      0.143
## Year2006      0.2636      0.1702      1.55      0.122
## Year2007      0.1474      0.1565      0.94      0.347
## Year2008     -0.1181      0.1551     -0.76      0.447
## Year2009     -0.0174      0.1505     -0.12      0.908
## Year2010      0.2041      0.1666      1.23      0.221
## Year2011      0.1509      0.1461      1.03      0.302
## Year2012      0.0524      0.1601      0.33      0.744
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.372, Adjusted R-squared:  0.341
## Convergence in 46 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 432 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0903 0.8580 0.9430 0.8820 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      2.14e-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 1.246 1      1.116
## LastAuthorFemale 1.203 1      1.097
## Year      1.324 16      1.009

```

Residuals from first and last author



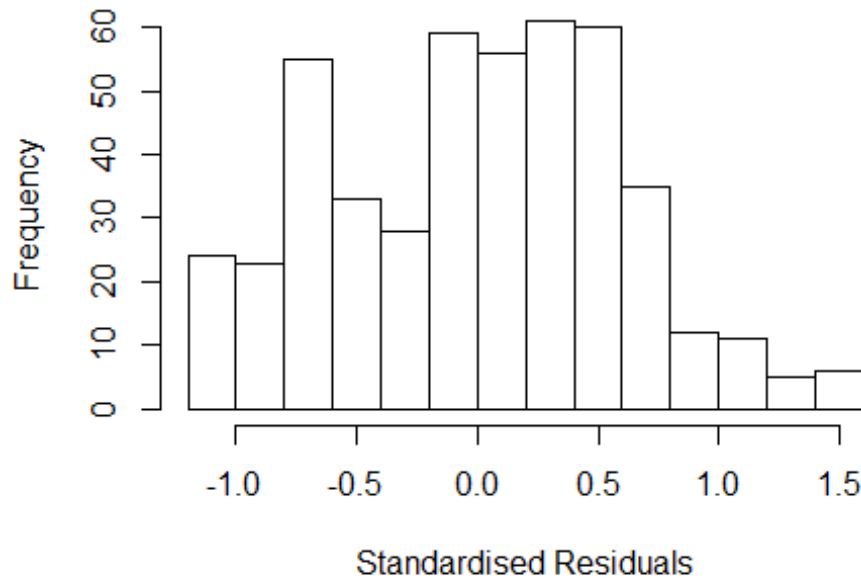
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1714 -0.4725 0.0338 0.4269 1.5183
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03095 0.13123 7.86 3e-14 ***
## FirstAuthorFemale1 0.14499 0.06357 2.28 0.0230 *
## LastAuthorFemale1 -0.17160 0.06324 -2.71 0.0069 **
## Year1997 -0.16490 0.18529 -0.89 0.3740
## Year1998 0.04663 0.30057 0.16 0.8768
## Year1999 0.06764 0.17681 0.38 0.7022
## Year2000 -0.02977 0.17530 -0.17 0.8652
## Year2001 0.00142 0.19089 0.01 0.9941
## Year2002 0.07175 0.18644 0.38 0.7005
## Year2003 0.10408 0.41155 0.25 0.8005
## Year2004 -0.09202 0.18069 -0.51 0.6108
## Year2005 0.07345 0.15865 0.46 0.6436
```

```

## Year2006          0.02907    0.16560    0.18    0.8607
## Year2007          0.01432    0.16507    0.09    0.9309
## Year2008         -0.37288    0.16364   -2.28    0.0232 *
## Year2009         -0.20779    0.15651   -1.33    0.1850
## Year2010          0.14049    0.15681    0.90    0.3708
## Year2011          0.06406    0.14500    0.44    0.6589
## Year2012         -0.12778    0.17009   -0.75    0.4529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.595
## Multiple R-squared:  0.0854, Adjusted R-squared:  0.0488
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 440 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.494  0.871  0.945  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      2.14e-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"
##           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
## -1.1386 -0.4724 0.0264 0.4250 1.5548
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08202 0.13670 7.92 1.9e-14 ***
## FirstAuthorFemale1 0.07622 0.06234 1.22 0.2221
## Year1997 -0.22768 0.19698 -1.16 0.2483
## Year1998 -0.05738 0.31283 -0.18 0.8546
## Year1999 -0.00315 0.18120 -0.02 0.9861
## Year2000 -0.08016 0.18500 -0.43 0.6650
## Year2001 -0.08539 0.18793 -0.45 0.6498
## Year2002 -0.05321 0.18335 -0.29 0.7718
## Year2003 -0.02340 0.36648 -0.06 0.9491
## Year2004 -0.17303 0.19076 -0.91 0.3649
## Year2005 -0.01129 0.16103 -0.07 0.9441
## Year2006 -0.08037 0.16614 -0.48 0.6288
```

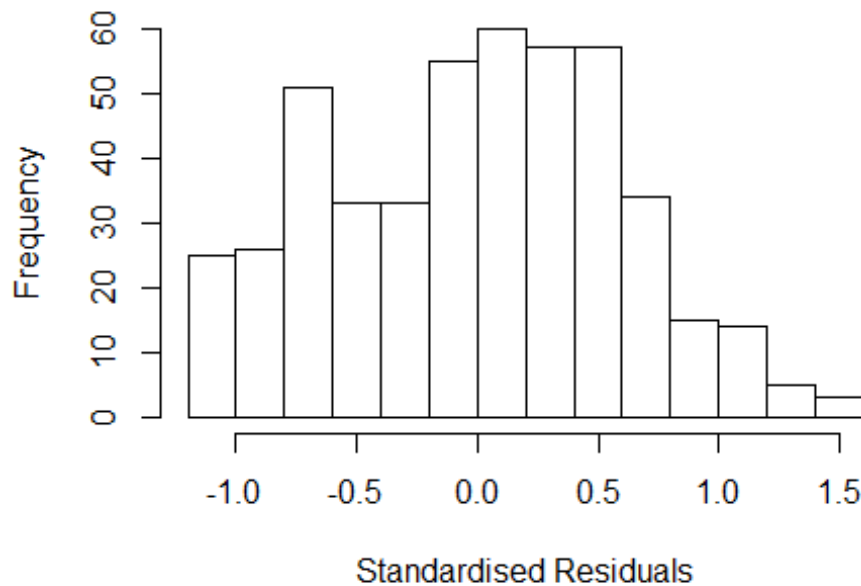


```

## Year2007          -0.07127    0.16887   -0.42    0.6732
## Year2008          -0.46544    0.16708   -2.79    0.0056 **
## Year2009          -0.31182    0.15889   -1.96    0.0503 .
## Year2010           0.05662    0.15824    0.36    0.7206
## Year2011          -0.02625    0.14775   -0.18    0.8591
## Year2012          -0.21754    0.17185   -1.27    0.2062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.0717, Adjusted R-squared:  0.0367
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 422 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.484  0.862  0.940  0.907  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      2.14e-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"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.162 1      1.078
## Year              1.162 16      1.005

```

Residuals from last author



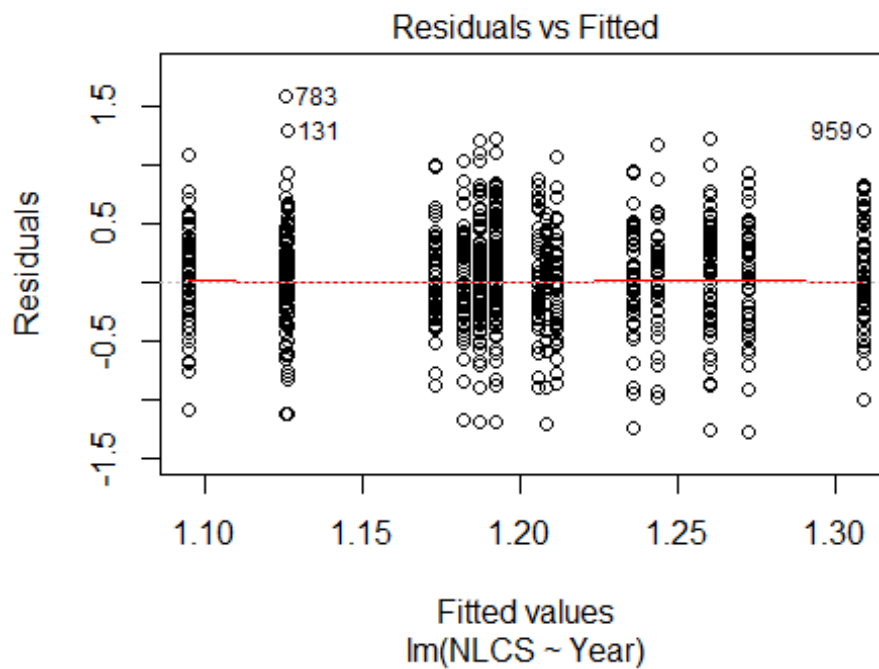
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.213 -0.494 0.037 0.414 1.584
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13875 0.13763 8.27 1.5e-15 ***
## LastAuthorFemale1 -0.11254 0.06232 -1.81 0.0716 .
## Year1997 -0.25826 0.19255 -1.34 0.1805
## Year1998 -0.02856 0.31728 -0.09 0.9283
## Year1999 -0.02357 0.18553 -0.13 0.8990
## Year2000 -0.07420 0.18626 -0.40 0.6906
## Year2001 -0.06311 0.19665 -0.32 0.7484
## Year2002 -0.02690 0.18922 -0.14 0.8870
## Year2003 0.06176 0.45411 0.14 0.8919
## Year2004 -0.14153 0.18940 -0.75 0.4553
## Year2005 0.00625 0.17045 0.04 0.9708
## Year2006 -0.04677 0.17312 -0.27 0.7872
```

```

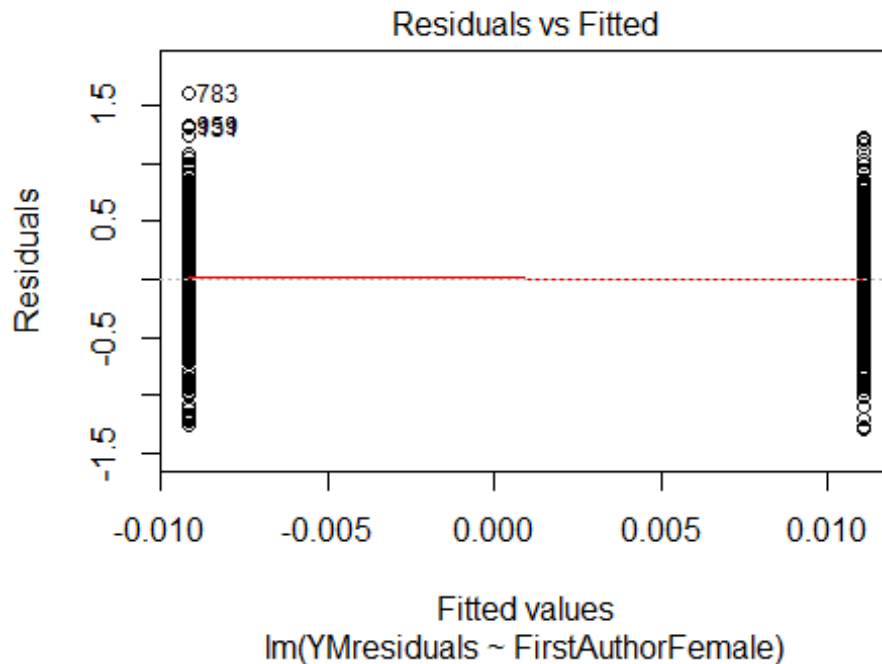
## Year2007          -0.07305      0.17169    -0.43    0.6707
## Year2008          -0.45022      0.17024    -2.64    0.0085 **
## Year2009          -0.27163      0.16954    -1.60    0.1098
## Year2010           0.07425      0.16795     0.44    0.6586
## Year2011           0.01255      0.15711     0.08    0.9364
## Year2012          -0.18184      0.18181    -1.00    0.3178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0773, Adjusted R-squared:  0.0424
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 429 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.458  0.865   0.939   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      2.14e-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: 468"
## [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
##   91  105   85   91   81   97   92   99   81   98   93   95   93   93   90
## 2011 2012
##  111  118
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   59   48   50   42   52   64   67   53   76   66   70   56   51   68
## 2011 2012

```

```
## 78 91
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 59 44 45 40 48 54 57 43 67 61 61 50 45 58
## 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 = 18, df = 16, p-value = 0.3
```

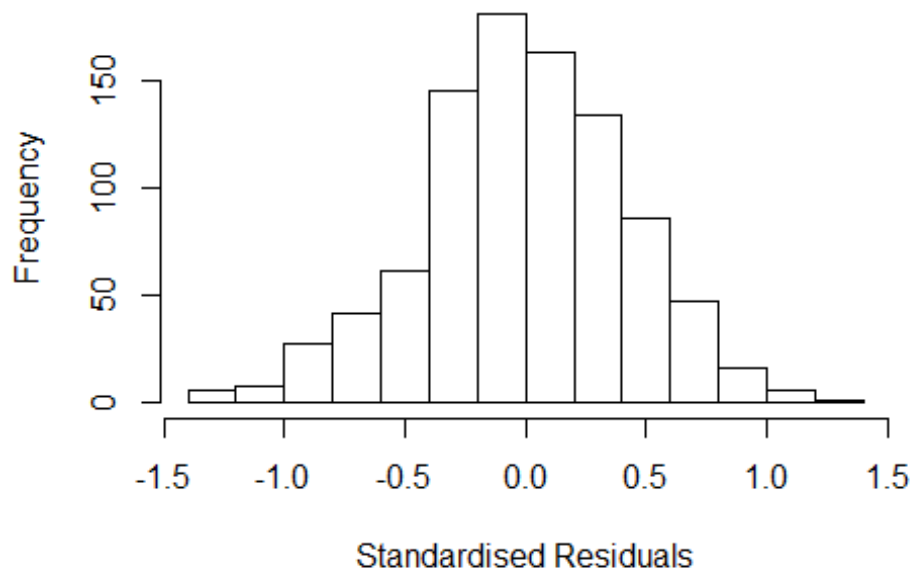


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.4, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 4.36355181392362"
## [1] "Male first author team size 2018 geometric mean: 4.59721091382268"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.25757654113476"
## [1] "Male last author team size 2018 geometric mean: 4.57487813639677"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, 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 1.176 1      1.084
## LastAuthorFemale  1.143 1      1.069
## UniqueAuthors    1.751 4      1.073
## Year              1.999 16     1.022
```

Residuals from first and last author and team size



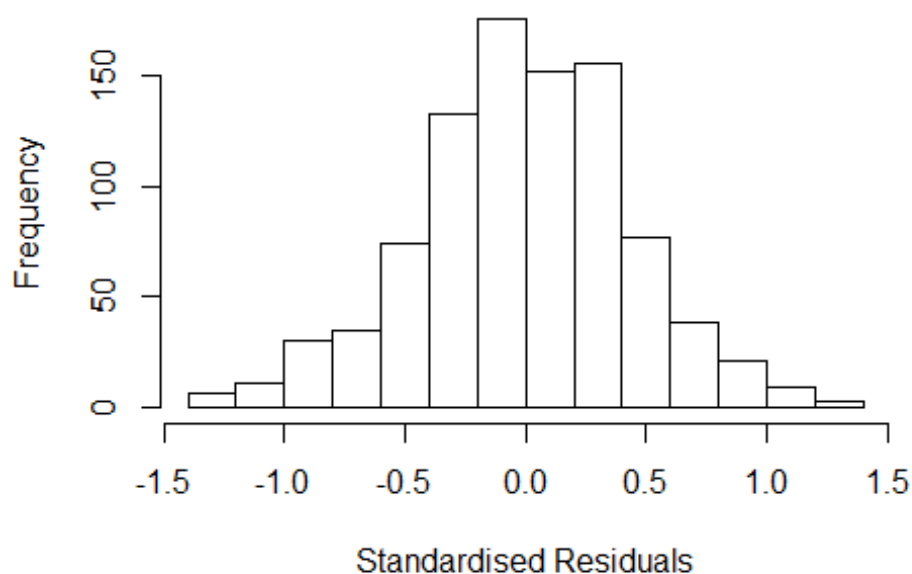
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3939 -0.2694 -0.0075  0.2747  1.2231
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05455    0.06906   15.27 < 2e-16 ***
## FirstAuthorFemale1 -0.00794    0.03022   -0.26  0.7929
## LastAuthorFemale1 -0.00744    0.03221   -0.23  0.8173
## UniqueAuthors2     0.13833    0.05698    2.43  0.0154 *
## UniqueAuthors3     0.24203    0.06211    3.90  0.0001 ***
## UniqueAuthors4     0.30763    0.06475    4.75 2.4e-06 ***
## UniqueAuthors5     0.36888    0.06189    5.96 3.6e-09 ***
## Year1997          -0.09065    0.09198   -0.99  0.3246
## Year1998          -0.01463    0.07455   -0.20  0.8445
## Year1999          -0.07513    0.08733   -0.86  0.3899
```

```

## Year2000      0.01533      0.10582      0.14      0.8848
## Year2001     -0.07681      0.08108     -0.95      0.3437
## Year2002     -0.12265      0.08196     -1.50      0.1349
## Year2003     -0.22256      0.07371     -3.02      0.0026 **
## Year2004     -0.12712      0.08614     -1.48      0.1404
## Year2005      0.00737      0.07860      0.09      0.9253
## Year2006     -0.10482      0.08275     -1.27      0.2056
## Year2007     -0.19334      0.08156     -2.37      0.0180 *
## Year2008     -0.02954      0.08743     -0.34      0.7355
## Year2009     -0.10303      0.08560     -1.20      0.2290
## Year2010     -0.04757      0.08547     -0.56      0.5779
## Year2011     -0.13142      0.08557     -1.54      0.1249
## Year2012     -0.01231      0.08472     -0.15      0.8845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0891, Adjusted R-squared:  0.0668
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 839 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.871  0.951  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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.135 1      1.065
## LastAuthorFemale  1.069 1      1.034
## Year              1.203 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
## -1.2985 -0.2742 -0.0062 0.2750 1.3264
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22018 0.06255 19.51 <2e-16 ***
## FirstAuthorFemale1 0.01442 0.03060 0.47 0.638
## LastAuthorFemale1 -0.00482 0.03189 -0.15 0.880
## Year1997 -0.08898 0.09699 -0.92 0.359
## Year1998 -0.00660 0.08089 -0.08 0.935
## Year1999 -0.06017 0.08693 -0.69 0.489
## Year2000 0.02891 0.10685 0.27 0.787
## Year2001 -0.05974 0.08238 -0.73 0.469
## Year2002 -0.07280 0.08291 -0.88 0.380
## Year2003 -0.16450 0.07706 -2.13 0.033 *
## Year2004 -0.08375 0.08682 -0.96 0.335
## Year2005 0.06847 0.08163 0.84 0.402
```

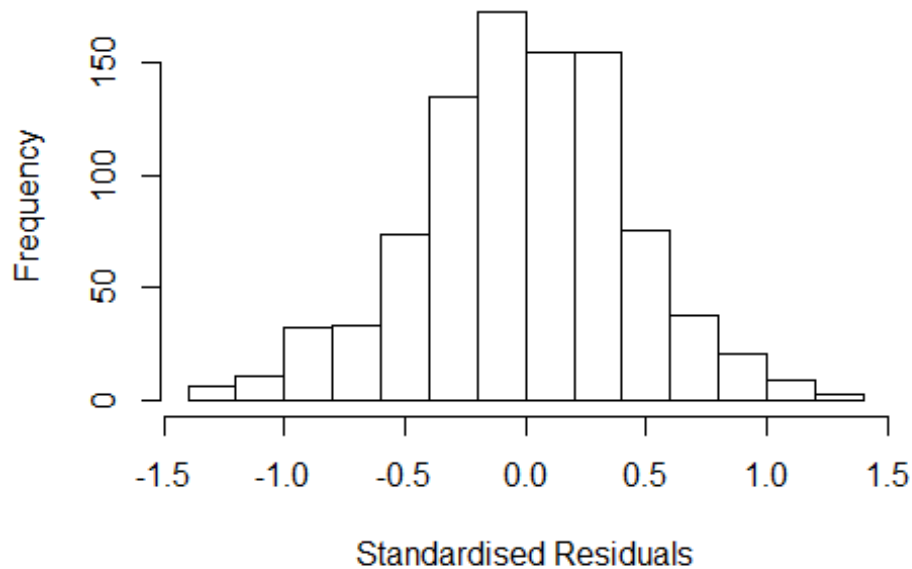


```

## Year2006      -0.03825    0.08411   -0.45    0.649
## Year2007      -0.12035    0.08388   -1.43    0.152
## Year2008       0.03816    0.09167    0.42    0.677
## Year2009      -0.04016    0.08749   -0.46    0.646
## Year2010       0.05204    0.08787    0.59    0.554
## Year2011      -0.03999    0.08900   -0.45    0.653
## Year2012       0.06386    0.08634    0.74    0.460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.00578
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 847 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.274  0.871  0.951  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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.128 1      1.062
## Year              1.128 16      1.004

```

Residuals from first author



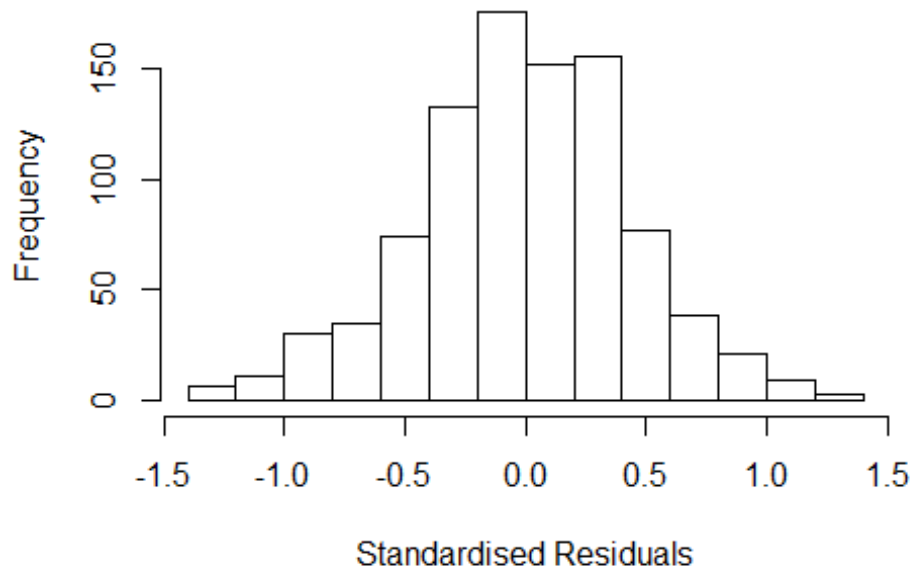
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2966 -0.2733 -0.0066  0.2754  1.3276
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.21905    0.06175   19.74  <2e-16 ***
## FirstAuthorFemale1  0.01388    0.03056    0.45   0.650
## Year1997         -0.08878    0.09702   -0.92   0.360
## Year1998         -0.00643    0.08088   -0.08   0.937
## Year1999         -0.06036    0.08693   -0.69   0.488
## Year2000          0.02853    0.10664    0.27   0.789
## Year2001         -0.05971    0.08229   -0.73   0.468
## Year2002         -0.07294    0.08287   -0.88   0.379
## Year2003         -0.16433    0.07705   -2.13   0.033 *
## Year2004         -0.08386    0.08686   -0.97   0.335
## Year2005          0.06835    0.08169    0.84   0.403
## Year2006         -0.03854    0.08420   -0.46   0.647
```

```

## Year2007      -0.12043    0.08391   -1.44    0.152
## Year2008      0.03833    0.09170    0.42    0.676
## Year2009     -0.04089    0.08715   -0.47    0.639
## Year2010      0.05227    0.08782    0.60    0.552
## Year2011     -0.04075    0.08882   -0.46    0.647
## Year2012      0.06366    0.08640    0.74    0.461
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.00687
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 851 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.272  0.870   0.952   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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.065 1          1.032
## Year              1.065 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
## -1.29184 -0.27675 -0.00604 0.27389 1.32012
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22393 0.06270 19.52 <2e-16 ***
## LastAuthorFemale1 -0.00309 0.03186 -0.10 0.923
## Year1997 -0.08815 0.09727 -0.91 0.365
## Year1998 -0.00682 0.08114 -0.08 0.933
## Year1999 -0.05948 0.08702 -0.68 0.494
## Year2000 0.03202 0.10663 0.30 0.764
## Year2001 -0.05922 0.08226 -0.72 0.472
## Year2002 -0.07313 0.08297 -0.88 0.378
## Year2003 -0.16080 0.07598 -2.12 0.035 *
## Year2004 -0.08004 0.08609 -0.93 0.353
## Year2005 0.07094 0.08071 0.88 0.380
## Year2006 -0.03716 0.08394 -0.44 0.658
```

```

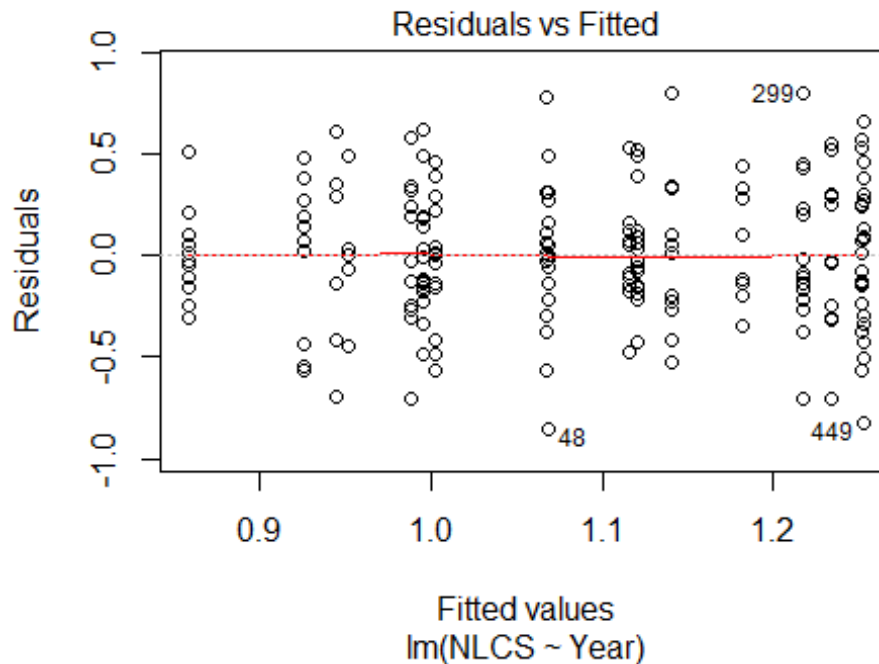
## Year2007          -0.11832      0.08381    -1.41      0.158
## Year2008           0.04317      0.09105      0.47      0.636
## Year2009          -0.03829      0.08729     -0.44      0.661
## Year2010           0.05422      0.08754      0.62      0.536
## Year2011          -0.03681      0.08891     -0.41      0.679
## Year2012           0.06791      0.08544      0.79      0.427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.025, Adjusted R-squared:  0.00667
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.277  0.873   0.953   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      1.09e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 921"
## [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]"
##
## 2001 2006 2007 2009 2011 2012
##    1    1    1    2    1    3
##
## 2001 2006 2007 2009 2011 2012
##    0    1    1    1    1    3
##
## 2001 2006 2007 2009 2011 2012
##    0    1    1    1    1    2

```

```

## [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: 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: 6"
## [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
## 21 31 35 32 28 22 42 20 24 14 22 28 21 23 17
## 2011 2012
## 18 13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 11 17 11 9 5 17 15 15 10 13 12 11 11 13
## 2011 2012
## 11 9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 8 14 9 8 4 15 11 12 9 11 9 10 9 13
## 2011 2012
## 11 9
## [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

```



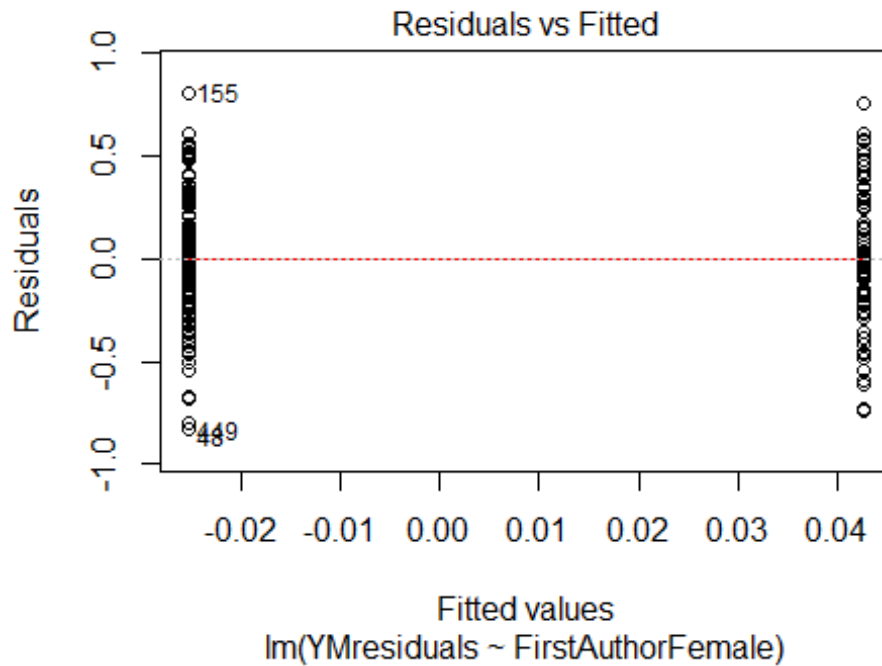
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.5, df = 1, p-value = 0.06

## [1] "Female first author team size 2018 geometric mean: 3.75694759634054"
## [1] "Male first author team size 2018 geometric mean: 5"

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

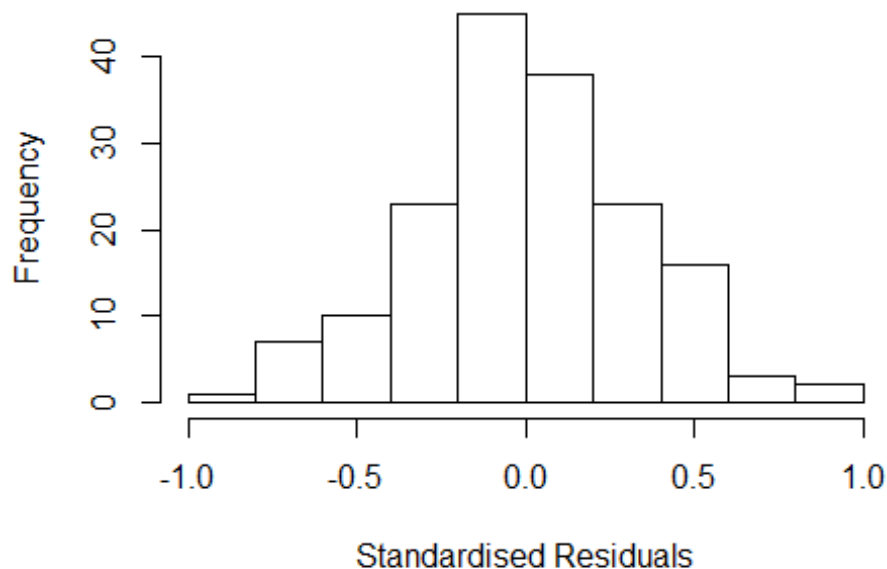
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2.5, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.62967808990371"
## [1] "Male last author team size 2018 geometric mean: 4.21286593061052"

## 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 = 8, 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 2.262 1      1.504
## LastAuthorFemale 2.238 1      1.496
## UniqueAuthors    7.982 4      1.296
## Year             20.485 16      1.099
```


Residuals from first and last author and team size



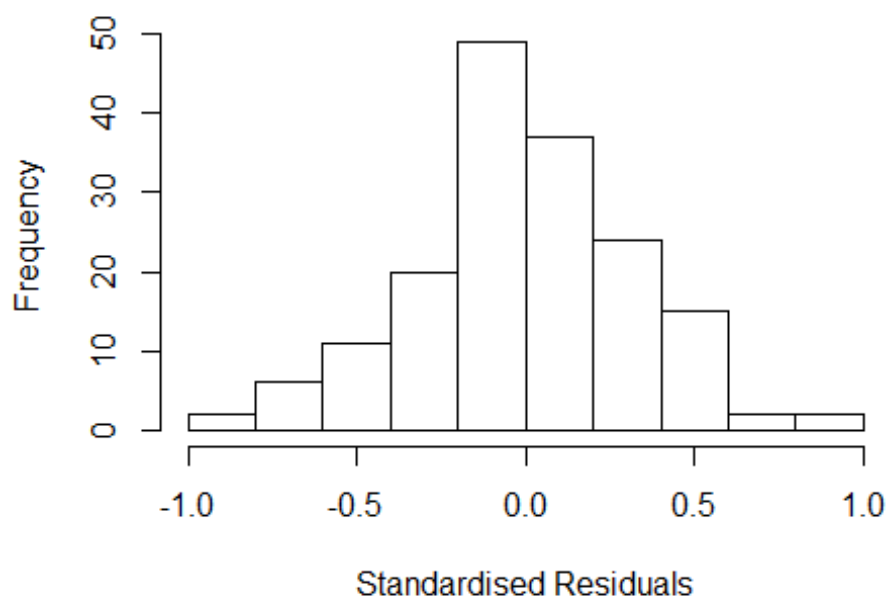
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.8472 -0.1915 -0.0076 0.2071 0.8807
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93489 0.33882 2.76 0.0065 **
## FirstAuthorFemale1 -0.07618 0.07898 -0.96 0.3364
## LastAuthorFemale1 0.00753 0.06736 0.11 0.9111
## UniqueAuthors2 0.03815 0.15043 0.25 0.8001
## UniqueAuthors3 0.04416 0.14503 0.30 0.7612
## UniqueAuthors4 -0.02037 0.14338 -0.14 0.8872
## UniqueAuthors5 0.07127 0.13437 0.53 0.5966
## Year1997 0.13772 0.31802 0.43 0.6656
## Year1998 0.21046 0.30985 0.68 0.4981
## Year1999 -0.14730 0.30235 -0.49 0.6269
```

```

## Year2000      0.19025    0.33270    0.57    0.5683
## Year2001      0.07781    0.35602    0.22    0.8273
## Year2002      0.10992    0.31026    0.35    0.7236
## Year2003      0.04973    0.30777    0.16    0.8718
## Year2004      0.21879    0.31835    0.69    0.4930
## Year2005      0.11645    0.34313    0.34    0.7348
## Year2006      0.28242    0.32090    0.88    0.3803
## Year2007      0.42374    0.33024    1.28    0.2015
## Year2008      0.15445    0.32401    0.48    0.6343
## Year2009      0.24703    0.30439    0.81    0.4184
## Year2010      0.34823    0.33552    1.04    0.3011
## Year2011      0.06166    0.31874    0.19    0.8469
## Year2012      0.26862    0.31737    0.85    0.3987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.308
## Multiple R-squared:  0.15, Adjusted R-squared:  0.0207
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 154 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.395  0.852  0.954  0.896  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      5.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.925 1      1.388
## LastAuthorFemale  2.073 1      1.440
## Year              3.081 16      1.036

```

Residuals from first and last author



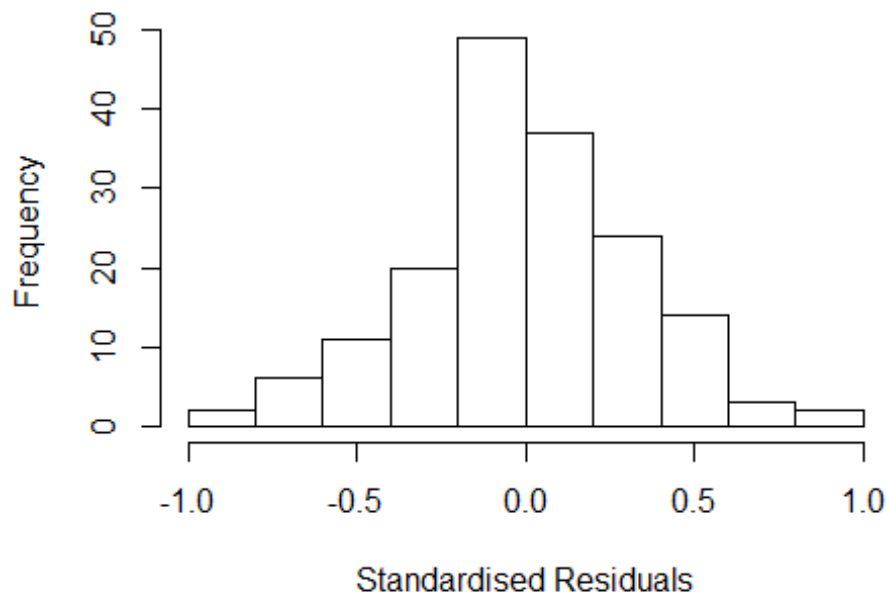
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.8135 -0.1883 -0.0114 0.2053 0.8143
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97342 0.27346 3.56 0.0005 ***
## FirstAuthorFemale1 -0.07079 0.07575 -0.93 0.3515
## LastAuthorFemale1 0.00543 0.06724 0.08 0.9358
## Year1997 0.12425 0.30577 0.41 0.6851
## Year1998 0.19726 0.28701 0.69 0.4930
## Year1999 -0.13821 0.28085 -0.49 0.6234
## Year2000 0.17742 0.31349 0.57 0.5723
## Year2001 0.06171 0.34313 0.18 0.8575
## Year2002 0.09373 0.28812 0.33 0.7454
## Year2003 0.05130 0.29307 0.18 0.8613
## Year2004 0.22630 0.30237 0.75 0.4554
## Year2005 0.11245 0.31283 0.36 0.7198
```

```

## Year2006          0.27736      0.29853      0.93      0.3543
## Year2007          0.42831      0.31151      1.37      0.1712
## Year2008          0.15136      0.31039      0.49      0.6265
## Year2009          0.24331      0.28534      0.85      0.3952
## Year2010          0.34187      0.30828      1.11      0.2692
## Year2011          0.05646      0.29934      0.19      0.8507
## Year2012          0.25282      0.29889      0.85      0.3990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.311
## Multiple R-squared:  0.14,   Adjusted R-squared:  0.0366
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 151 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.473  0.851  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      5.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.612 1      1.270
## Year              1.612 16      1.015

```

Residuals from first author



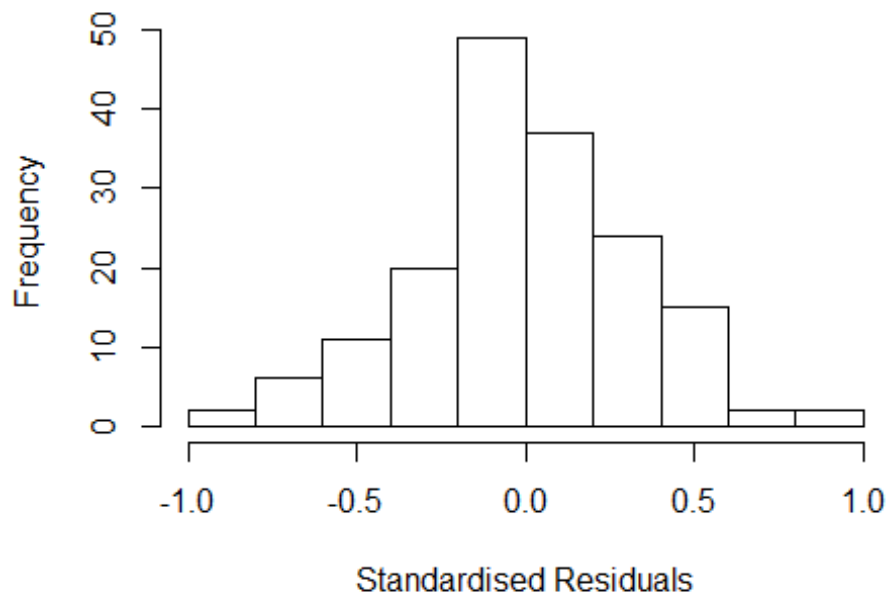
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.8142 -0.1884 -0.0106 0.2061 0.8101
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9708 0.2658 3.65 0.00036 ***
## FirstAuthorFemale1 -0.0694 0.0698 -0.99 0.32191
## Year1997 0.1279 0.2920 0.44 0.66204
## Year1998 0.2011 0.2805 0.72 0.47446
## Year1999 -0.1348 0.2736 -0.49 0.62298
## Year2000 0.1824 0.3054 0.60 0.55130
## Year2001 0.0675 0.3314 0.20 0.83884
## Year2002 0.0961 0.2808 0.34 0.73253
## Year2003 0.0547 0.2852 0.19 0.84812
## Year2004 0.2331 0.2928 0.80 0.42734
## Year2005 0.1127 0.3062 0.37 0.71324
## Year2006 0.2804 0.2912 0.96 0.33707
```

```

## Year2007          0.4288      0.3048      1.41  0.16154
## Year2008          0.1577      0.3025      0.52  0.60286
## Year2009          0.2477      0.2780      0.89  0.37433
## Year2010          0.3438      0.2998      1.15  0.25343
## Year2011          0.0593      0.2931      0.20  0.84007
## Year2012          0.2559      0.2915      0.88  0.38142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.32
## Multiple R-squared:  0.138, Adjusted R-squared:  0.0405
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 149 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.497  0.856  0.951  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      5.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.687 1          1.299
## Year            1.687 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
## -0.85911 -0.19571  0.00325  0.23105  0.86393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97062    0.26500   3.66  0.00035 ***
## LastAuthorFemale1 -0.01291    0.06151  -0.21  0.83406
## Year1997         0.09240    0.29786   0.31  0.75682
## Year1998         0.16244    0.27698   0.59  0.55844
## Year1999        -0.17689    0.27085  -0.65  0.51468
## Year2000         0.13168    0.30010   0.44  0.66144
## Year2001         0.00919    0.32919   0.03  0.97777
## Year2002         0.04941    0.27463   0.18  0.85747
## Year2003         0.01994    0.28402   0.07  0.94413
## Year2004         0.19949    0.29177   0.68  0.49521
## Year2005         0.05122    0.28733   0.18  0.85877
## Year2006         0.22492    0.28206   0.80  0.42648
```

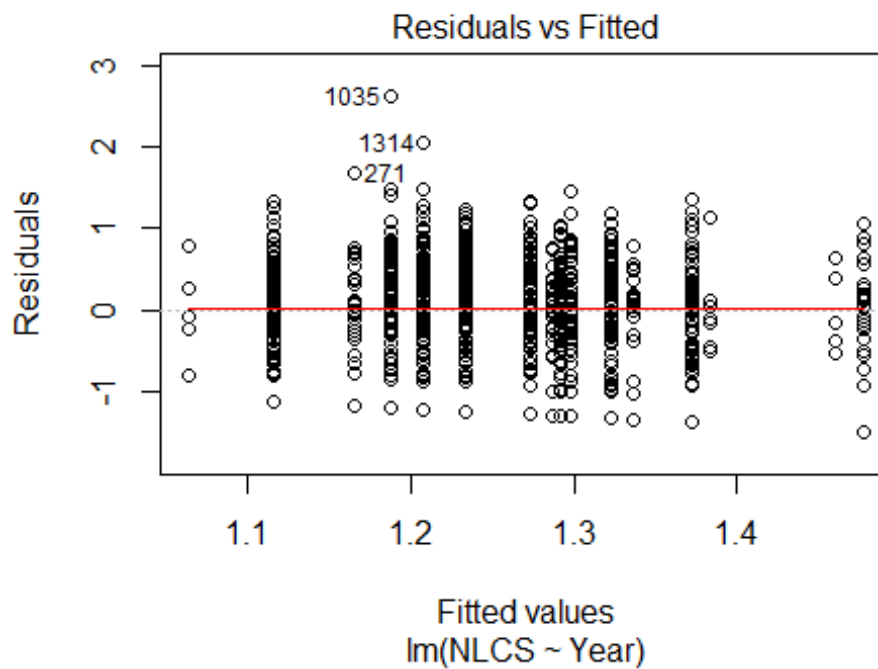
```

## Year2007      0.37838    0.30442    1.24  0.21582
## Year2008      0.10045    0.29438    0.34  0.73342
## Year2009      0.20808    0.27456    0.76  0.44973
## Year2010      0.31949    0.29960    1.07  0.28796
## Year2011      0.03279    0.29012    0.11  0.91015
## Year2012      0.20902    0.28719    0.73  0.46788
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.131, Adjusted R-squared:  0.0326
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 150 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.448  0.847  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      5.95e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 168"
## [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
##   24   18   20   22   35   47   39   37   68   75  108  132  160  142  160
## 2011 2012
##  183  191
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    5    8    8   21   26   26   30   57   62   79  115  138  111  124
## 2011 2012

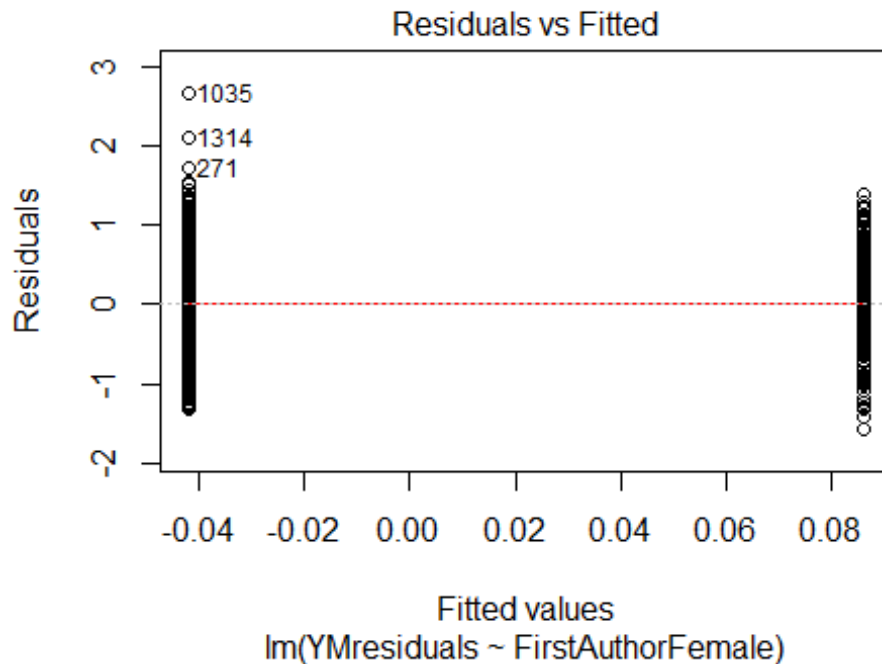
```



```
## 157 142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 5 6 7 20 24 23 25 49 55 68 106 120 98 111
## 2011 2012
## 140 133
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 8.2, df = 16, p-value = 0.9
```

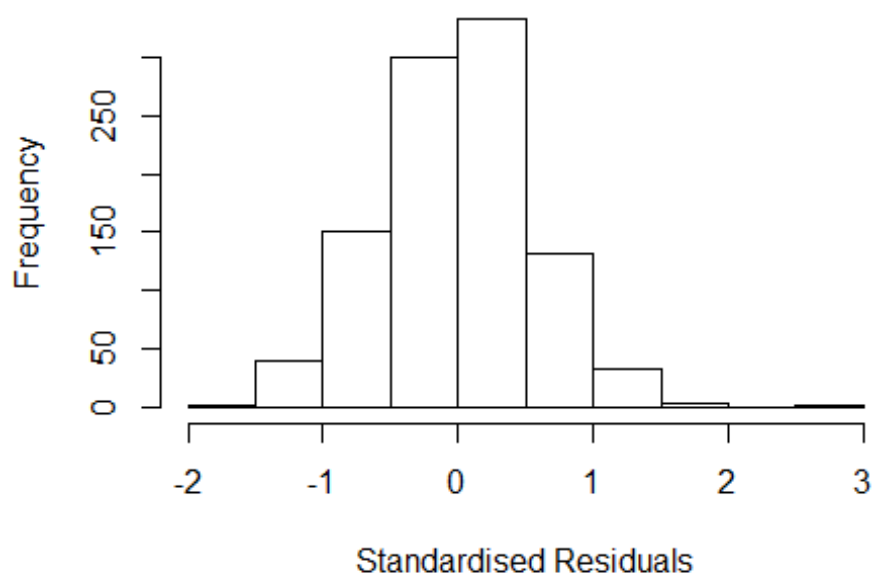


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7, df = 1, p-value = 0.008
```



```
## [1] "Female first author team size 2018 geometric mean: 3.78328053938118"
## [1] "Male first author team size 2018 geometric mean: 4.28963265911291"
##
## 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: 3.50849986412189"
## [1] "Male last author team size 2018 geometric mean: 4.27813775999801"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, 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 1.129 1      1.062
## LastAuthorFemale  1.108 1      1.053
## UniqueAuthors    1.556 4      1.057
## Year             1.736 16      1.017
```

Residuals from first and last author and team size



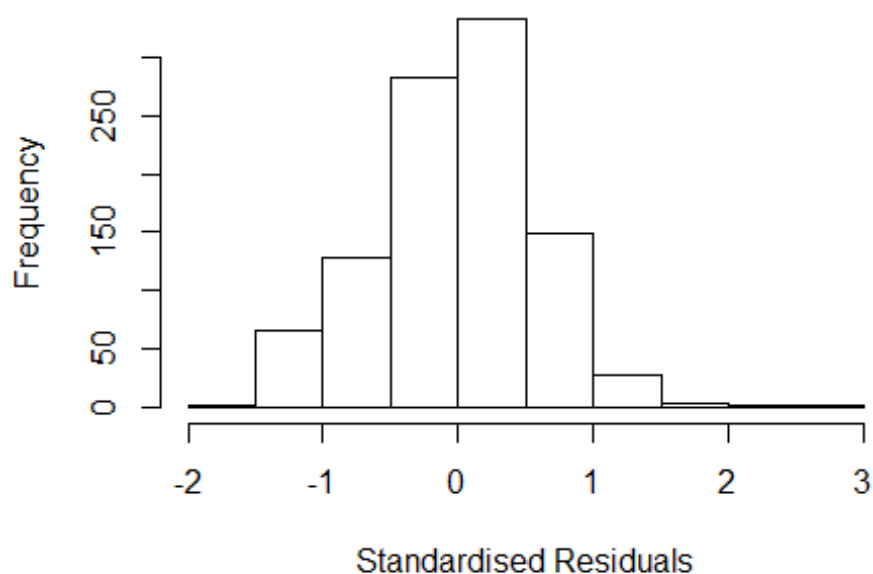
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1035 78649579593 3.803 2010      2711      2      2.715
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.53137 -0.33856  0.00544  0.35726  2.71487
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5515     0.2628   5.90 4.9e-09 ***
## FirstAuthorFemale1  0.0839     0.0376   2.23 0.02586 *
## LastAuthorFemale1  0.0763     0.0397   1.92 0.05532 .
## UniqueAuthors2     0.3048     0.0767   3.97 7.6e-05 ***
## UniqueAuthors3     0.4492     0.0725   6.19 8.8e-10 ***
## UniqueAuthors4     0.5262     0.0764   6.89 1.0e-11 ***
## UniqueAuthors5     0.6551     0.0705   9.29 < 2e-16 ***
## Year1997          -0.8519     0.3974  -2.14 0.03228 *
## Year1998          -0.6364     0.3874  -1.64 0.10073
## Year1999          -0.6910     0.3476  -1.99 0.04711 *
```

```

## Year2000          -0.5301      0.2860    -1.85   0.06413 .
## Year2001          -0.5473      0.2872    -1.91   0.05697 .
## Year2002          -0.6460      0.2742    -2.36   0.01866 *
## Year2003          -0.8885      0.2839    -3.13   0.00180 **
## Year2004          -0.7208      0.2758    -2.61   0.00911 **
## Year2005          -0.6752      0.2754    -2.45   0.01438 *
## Year2006          -0.6433      0.2703    -2.38   0.01750 *
## Year2007          -0.7706      0.2663    -2.89   0.00389 **
## Year2008          -0.9270      0.2690    -3.45   0.00059 ***
## Year2009          -0.7515      0.2697    -2.79   0.00543 **
## Year2010          -0.9125      0.2685    -3.40   0.00070 ***
## Year2011          -0.8490      0.2672    -3.18   0.00153 **
## Year2012          -0.8774      0.2692    -3.26   0.00116 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.152, Adjusted R-squared:  0.133
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 611 is an outlier with |weight| = 0 ( < 0.0001);
## 84 weights are ~= 1. The remaining 908 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.166  0.865  0.950  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           1.01e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.096 1 1.047
## LastAuthorFemale 1.086 1 1.042
## Year 1.187 16 1.005

```

Residuals from first and last author



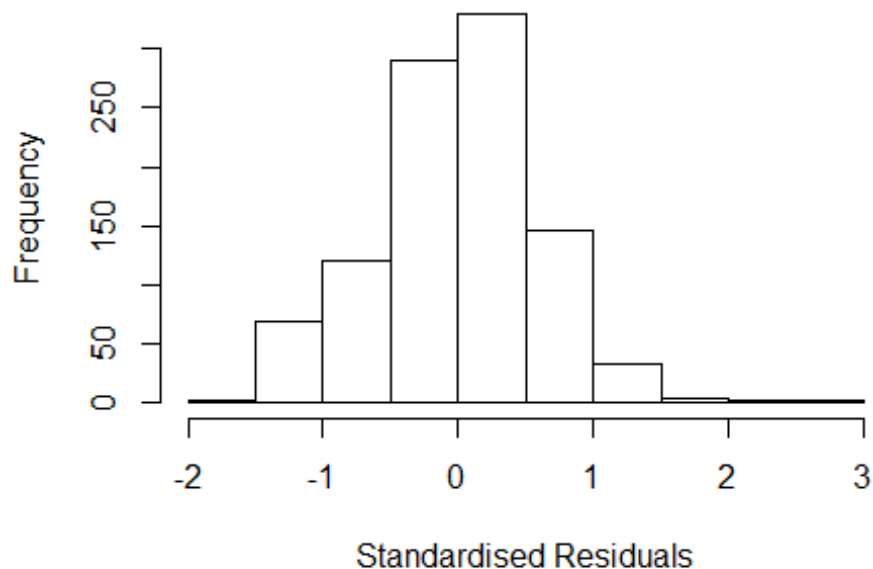
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1035 78649579593 3.803 2010      2711      2      2.714
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5120 -0.3864  0.0194  0.3717  2.7139
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6582     0.3222   5.15 3.2e-07 ***
## FirstAuthorFemale1  0.1206     0.0393   3.07  0.0022 **
## LastAuthorFemale1  0.0865     0.0418   2.07  0.0388 *
## Year1997          -0.6556     0.4116  -1.59  0.1115
## Year1998          -0.3417     0.3822  -0.89  0.3715
## Year1999          -0.5072     0.3698  -1.37  0.1705
## Year2000          -0.3293     0.3411  -0.97  0.3347
## Year2001          -0.2668     0.3429  -0.78  0.4366
## Year2002          -0.3363     0.3325  -1.01  0.3120
## Year2003          -0.6785     0.3455  -1.96  0.0498 *
## Year2004          -0.4846     0.3334  -1.45  0.1464
## Year2005          -0.4211     0.3357  -1.25  0.2100
```

```

## Year2006          -0.3237      0.3296   -0.98   0.3264
## Year2007          -0.4241      0.3270   -1.30   0.1949
## Year2008          -0.5924      0.3277   -1.81   0.0709 .
## Year2009          -0.3954      0.3282   -1.20   0.2287
## Year2010          -0.5691      0.3290   -1.73   0.0840 .
## Year2011          -0.5020      0.3268   -1.54   0.1249
## Year2012          -0.4897      0.3279   -1.49   0.1356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0431, Adjusted R-squared:  0.0254
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 611 is an outlier with |weight| = 0 ( < 0.0001);
## 80 weights are ~= 1. The remaining 912 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.126  0.859   0.950   0.901   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           1.01e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## 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



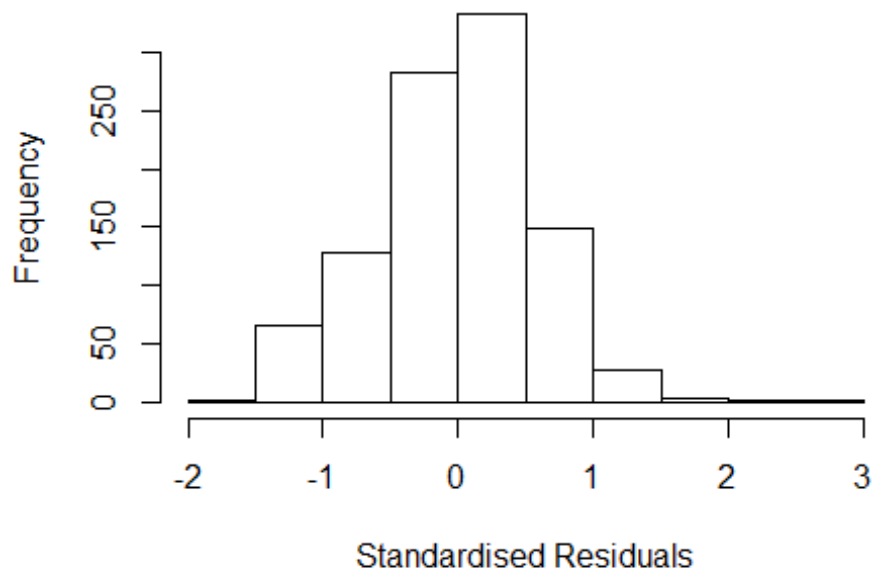
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1035 78649579593 3.803 2010      2711      2      2.714
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5644 -0.3780  0.0288  0.3631  2.6912
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6583     0.3222   5.15 3.2e-07 ***
## FirstAuthorFemale1 0.1307     0.0394   3.31 0.00096 ***
## Year1997        -0.6605     0.4116  -1.60 0.10890
## Year1998        -0.3425     0.3868  -0.89 0.37609
## Year1999        -0.4996     0.3737  -1.34 0.18159
## Year2000        -0.2974     0.3403  -0.87 0.38234
## Year2001        -0.2245     0.3422  -0.66 0.51191
## Year2002        -0.3162     0.3327  -0.95 0.34216
## Year2003        -0.6495     0.3447  -1.88 0.05987 .
## Year2004        -0.4644     0.3332  -1.39 0.16372
## Year2005        -0.3998     0.3355  -1.19 0.23374
## Year2006        -0.2905     0.3292  -0.88 0.37776
```

```

## Year2007          -0.4093      0.3270   -1.25   0.21104
## Year2008          -0.5661      0.3274   -1.73   0.08418 .
## Year2009          -0.3754      0.3283   -1.14   0.25312
## Year2010          -0.5465      0.3286   -1.66   0.09663 .
## Year2011          -0.4809      0.3266   -1.47   0.14126
## Year2012          -0.4695      0.3277   -1.43   0.15228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0384, Adjusted R-squared:  0.0216
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 611 is an outlier with |weight| = 0 ( < 0.0001);
## 78 weights are ~= 1. The remaining 914 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.134  0.859  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      1.01e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.089 1          1.043
## Year            1.089 16          1.003

```


Residuals from last author



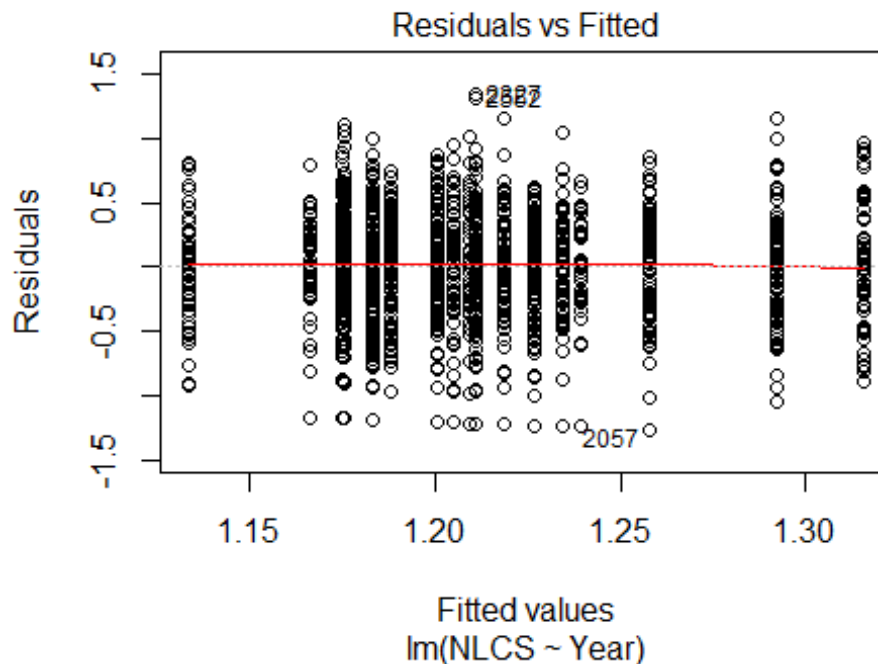
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1035 78649579593 3.803 2010      2711      2      2.714
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4161 -0.3618  0.0424  0.3655  2.6689
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6583     0.3223   5.15 3.2e-07 ***
## LastAuthorFemale1  0.1020     0.0419   2.43  0.015 *
## Year1997        -0.5970     0.4112  -1.45  0.147
## Year1998        -0.2920     0.3808  -0.77  0.443
## Year1999        -0.4770     0.3689  -1.29  0.196
## Year2000        -0.2932     0.3409  -0.86  0.390
## Year2001        -0.2422     0.3426  -0.71  0.480
## Year2002        -0.3112     0.3327  -0.94  0.350
## Year2003        -0.6498     0.3442  -1.89  0.059 .
## Year2004        -0.4389     0.3334  -1.32  0.188
## Year2005        -0.3792     0.3353  -1.13  0.258
## Year2006        -0.2858     0.3293  -0.87  0.386
```

```

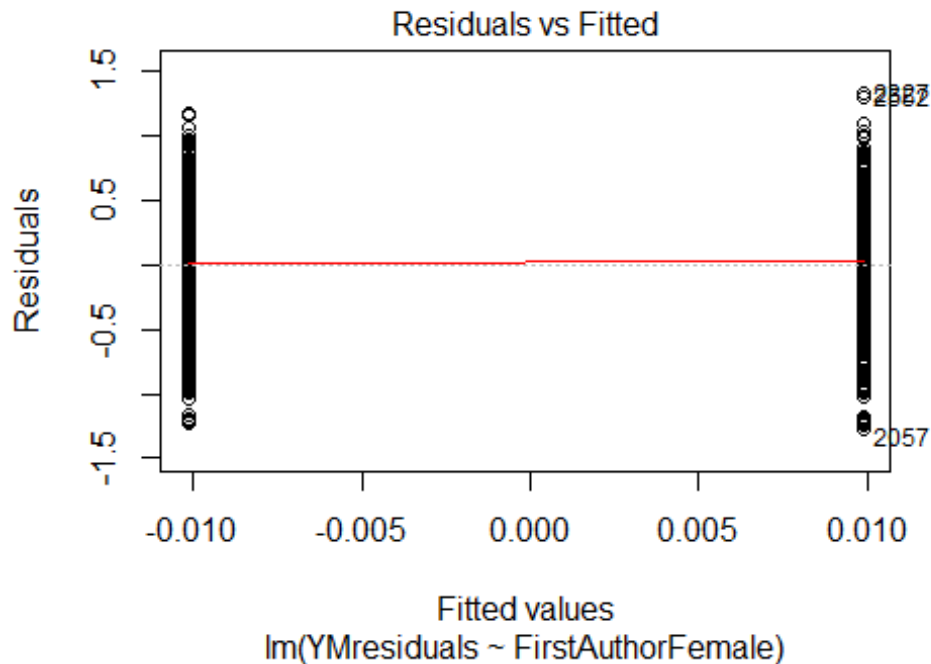
## Year2007          -0.3875      0.3271    -1.18      0.236
## Year2008          -0.5647      0.3277    -1.72      0.085 .
## Year2009          -0.3587      0.3280    -1.09      0.274
## Year2010          -0.5242      0.3286    -1.60      0.111
## Year2011          -0.4667      0.3268    -1.43      0.154
## Year2012          -0.4385      0.3270    -1.34      0.180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0334, Adjusted R-squared:  0.0166
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 611 is an outlier with |weight| = 0 ( < 0.0001);
## 69 weights are ~= 1. The remaining 923 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.140  0.857  0.953  0.901  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      1.01e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 993"
## [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
## 154 141 102 109 144 167 147 124 143 184 179 194 216 240 250
## 2011 2012
## 266 279
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 59 41 50 50 63 84 81 80 113 118 103 128 133 139

```

```
## 2011 2012
## 135 154
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 64 54 34 45 40 56 76 72 69 92 107 93 111 116 123
## 2011 2012
## 120 140
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 16, df = 16, p-value = 0.5
```

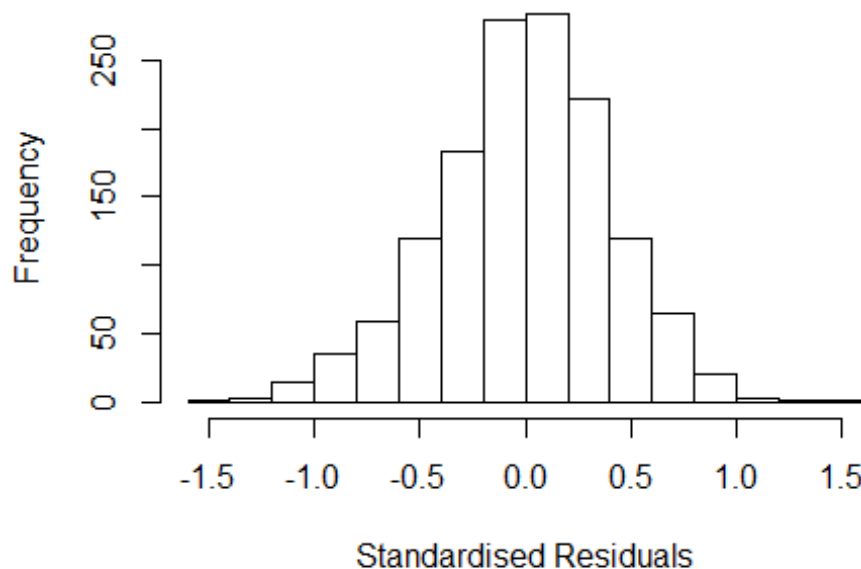


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 5.5801010462101"
## [1] "Male first author team size 2018 geometric mean: 4.97879709469479"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.30354826746803"
## [1] "Male last author team size 2018 geometric mean: 5.38875243962648"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, 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 1.101 1      1.049
## LastAuthorFemale  1.042 1      1.021
## UniqueAuthors    1.436 4      1.046
## Year              1.475 16     1.012
```

Residuals from first and last author and team size



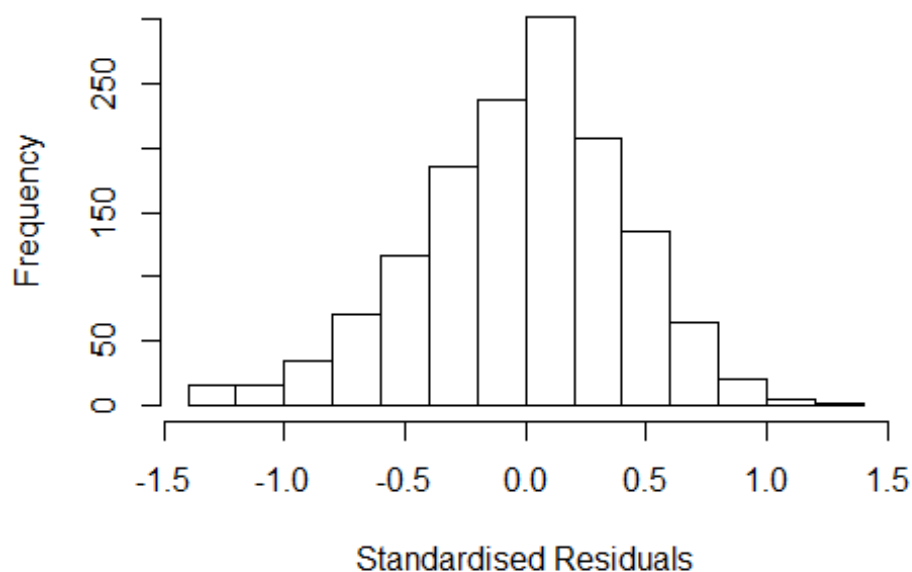
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40842 -0.26392  0.00833  0.26462  1.49540
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9619    0.1006   9.56  <2e-16 ***
## FirstAuthorFemale1  0.0109    0.0230   0.48  0.6337
## LastAuthorFemale1 -0.0337    0.0261  -1.29  0.1976
## UniqueAuthors2    0.1133    0.0945   1.20  0.2304
## UniqueAuthors3    0.2736    0.0917   2.98  0.0029 **
## UniqueAuthors4    0.2865    0.0914   3.13  0.0018 **
## UniqueAuthors5    0.4465    0.0884   5.05  5e-07 ***
## Year1997         0.0572    0.1000   0.57  0.5674
## Year1998        -0.0423    0.0816  -0.52  0.6048
## Year1999        -0.0302    0.0942  -0.32  0.7487
```

```

## Year2000          -0.1385      0.0871   -1.59   0.1119
## Year2001          -0.1351      0.0797   -1.70   0.0900 .
## Year2002           0.0352      0.0735    0.48   0.6319
## Year2003          -0.0627      0.0717   -0.88   0.3816
## Year2004          -0.0682      0.0760   -0.90   0.3697
## Year2005          -0.1043      0.0684   -1.52   0.1279
## Year2006          -0.0499      0.0676   -0.74   0.4601
## Year2007          -0.0235      0.0697   -0.34   0.7359
## Year2008          -0.1581      0.0695   -2.27   0.0231 *
## Year2009          -0.0653      0.0683   -0.96   0.3390
## Year2010          -0.1446      0.0674   -2.14   0.0322 *
## Year2011          -0.0940      0.0706   -1.33   0.1830
## Year2012          -0.1311      0.0664   -1.97   0.0486 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.108, Adjusted R-squared:  0.094
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1294 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.120  0.866  0.949  0.901  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          7.08e-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 1.060 1          1.030
## LastAuthorFemale  1.031 1          1.015
## Year              1.092 16          1.003

```

Residuals from first and last author



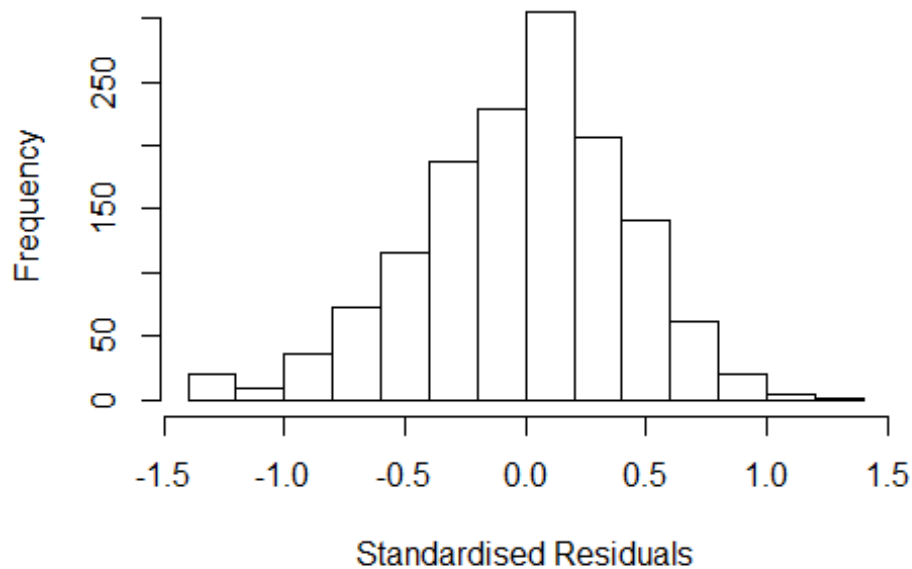
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2797 -0.2773  0.0167  0.2656  1.3976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2365     0.0559   22.10  <2e-16 ***
## FirstAuthorFemale1  0.0244     0.0236    1.03   0.301
## LastAuthorFemale1 -0.0462     0.0274   -1.68   0.092 .
## Year1997          0.0554     0.0985    0.56   0.574
## Year1998          0.0600     0.0807    0.74   0.458
## Year1999         -0.0384     0.0898   -0.43   0.669
## Year2000         -0.0791     0.0843   -0.94   0.349
## Year2001         -0.0945     0.0803   -1.18   0.240
## Year2002          0.0758     0.0726    1.05   0.296
## Year2003          0.0356     0.0713    0.50   0.617
## Year2004          0.0119     0.0766    0.16   0.877
## Year2005         -0.0514     0.0698   -0.74   0.462
```

```

## Year2006          0.0301      0.0665      0.45      0.651
## Year2007          0.0618      0.0681      0.91      0.364
## Year2008         -0.0624      0.0686     -0.91      0.363
## Year2009          0.0188      0.0689      0.27      0.785
## Year2010         -0.0628      0.0674     -0.93      0.351
## Year2011         -0.0374      0.0703     -0.53      0.595
## Year2012         -0.0391      0.0655     -0.60      0.551
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00402
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1288 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.217  0.860  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      7.08e-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 1.058 1      1.028
## Year              1.058 16      1.002

```


Residuals from first author



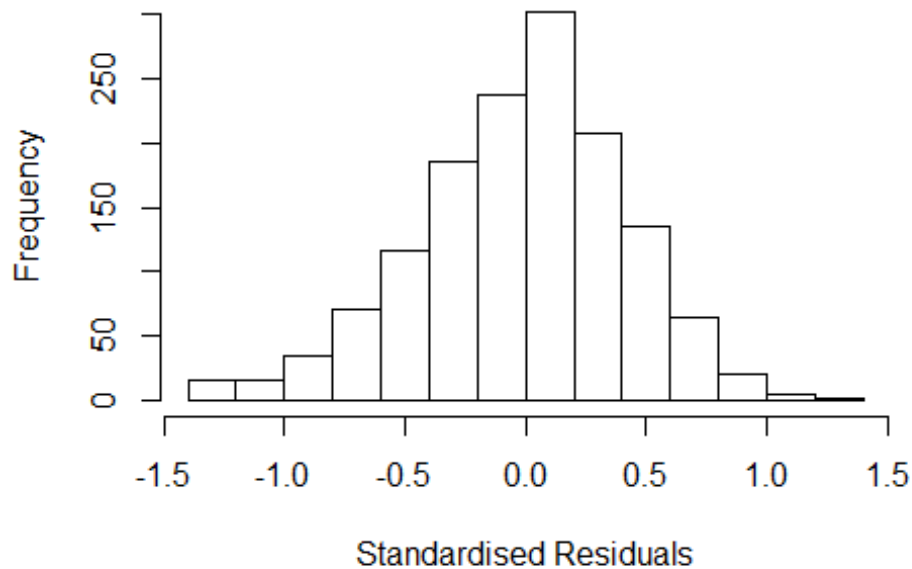
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.310 -0.281 0.022 0.269 1.363
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2316 0.0557 22.13 <2e-16 ***
## FirstAuthorFemale1 0.0221 0.0236 0.94 0.35
## Year1997 0.0532 0.0978 0.54 0.59
## Year1998 0.0559 0.0817 0.68 0.49
## Year1999 -0.0400 0.0895 -0.45 0.66
## Year2000 -0.0764 0.0841 -0.91 0.36
## Year2001 -0.0946 0.0803 -1.18 0.24
## Year2002 0.0729 0.0724 1.01 0.31
## Year2003 0.0340 0.0712 0.48 0.63
## Year2004 0.0104 0.0763 0.14 0.89
## Year2005 -0.0530 0.0695 -0.76 0.45
## Year2006 0.0245 0.0663 0.37 0.71
```

```

## Year2007          0.0561      0.0681      0.83      0.41
## Year2008          -0.0670      0.0685     -0.98      0.33
## Year2009           0.0121      0.0688      0.18      0.86
## Year2010          -0.0686      0.0676     -1.02      0.31
## Year2011          -0.0465      0.0701     -0.66      0.51
## Year2012          -0.0499      0.0651     -0.77      0.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.00261
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1291 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.241  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      7.08e-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 1.031 1          1.016
## Year            1.031 16          1.001

```

Residuals from last author



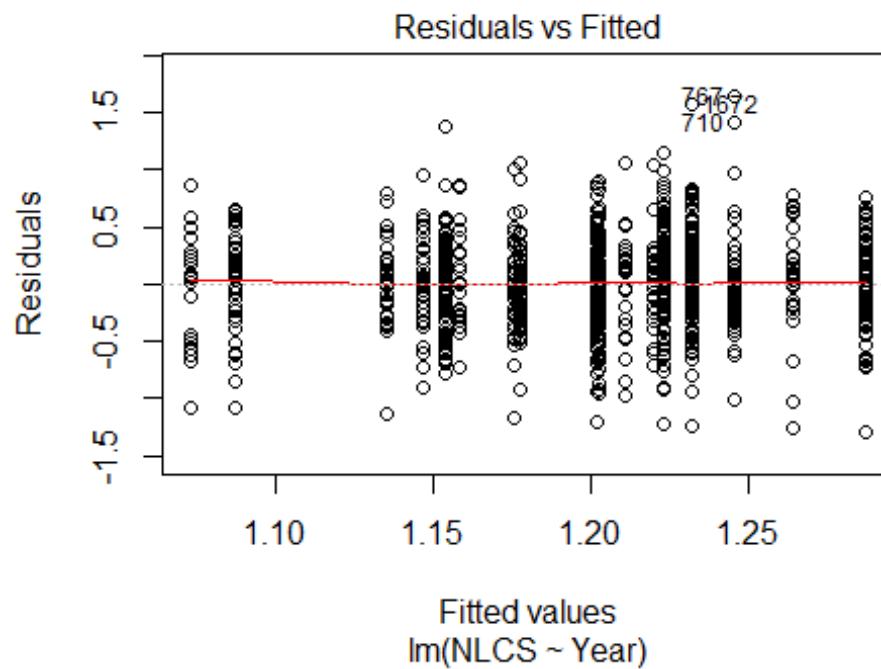
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2829 -0.2760 0.0191 0.2720 1.4061
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2456 0.0550 22.67 <2e-16 ***
## LastAuthorFemale1 -0.0444 0.0274 -1.62 0.11
## Year1997 0.0540 0.0983 0.55 0.58
## Year1998 0.0578 0.0811 0.71 0.48
## Year1999 -0.0351 0.0896 -0.39 0.70
## Year2000 -0.0777 0.0847 -0.92 0.36
## Year2001 -0.0925 0.0806 -1.15 0.25
## Year2002 0.0791 0.0726 1.09 0.28
## Year2003 0.0374 0.0713 0.52 0.60
## Year2004 0.0171 0.0766 0.22 0.82
## Year2005 -0.0489 0.0700 -0.70 0.48
## Year2006 0.0326 0.0667 0.49 0.63
```

```

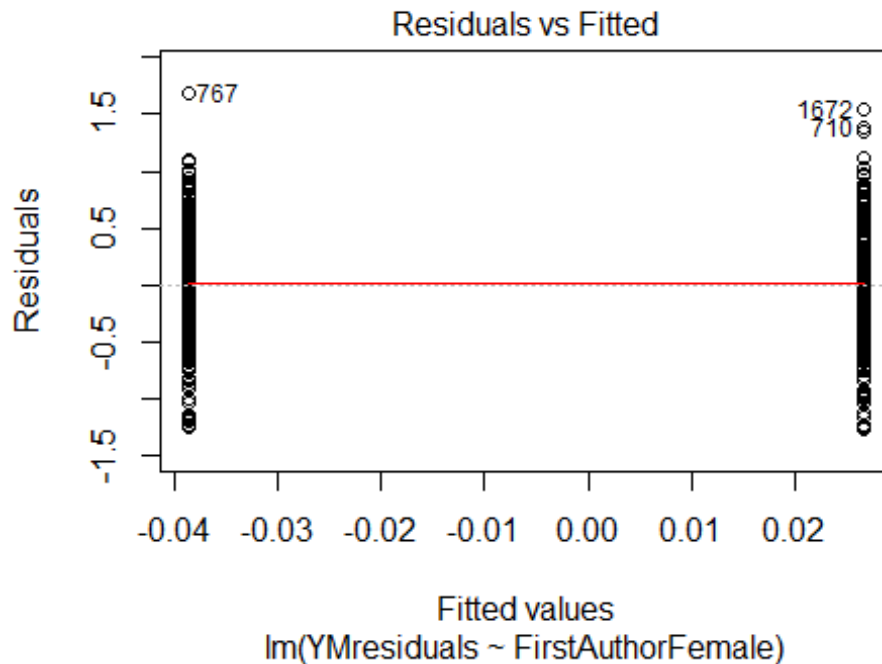
## Year2007          0.0678      0.0680      1.00      0.32
## Year2008          -0.0572     0.0686     -0.83      0.40
## Year2009           0.0223     0.0691      0.32      0.75
## Year2010          -0.0585     0.0675     -0.87      0.39
## Year2011          -0.0336     0.0703     -0.48      0.63
## Year2012          -0.0330     0.0654     -0.50      0.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00401
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1295 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.210  0.863  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      7.08e-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: 1412"
## [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
##   54   40   51   56   50   64   70   74   60   78   99   95  106  125  123
## 2011 2012
##  167  176
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   24   29   36   22   27   45   41   45   62   73   73   81   84   96
## 2011 2012

```

```
## 115 133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 21 27 29 19 21 38 34 40 54 60 66 71 78 85
## 2011 2012
## 104 121
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.2
```

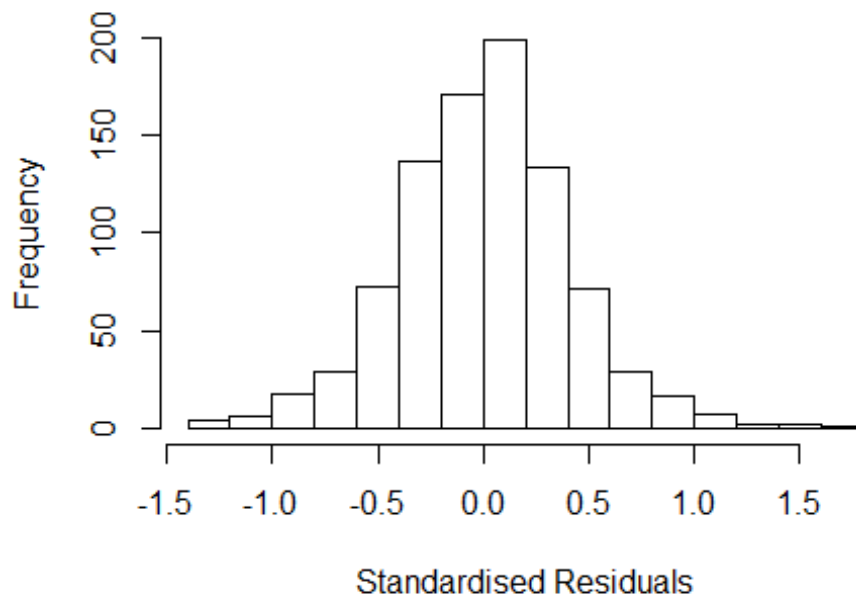


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.77, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 5.47009596055622"
## [1] "Male first author team size 2018 geometric mean: 4.28217462744037"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.13199865632105"
## [1] "Male last author team size 2018 geometric mean: 4.75804910424762"
##
## 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 1.154 1      1.074
## LastAuthorFemale  1.142 1      1.068
## UniqueAuthors    1.673 4      1.066
## Year              1.791 16     1.018
```

Residuals from first and last author and team size



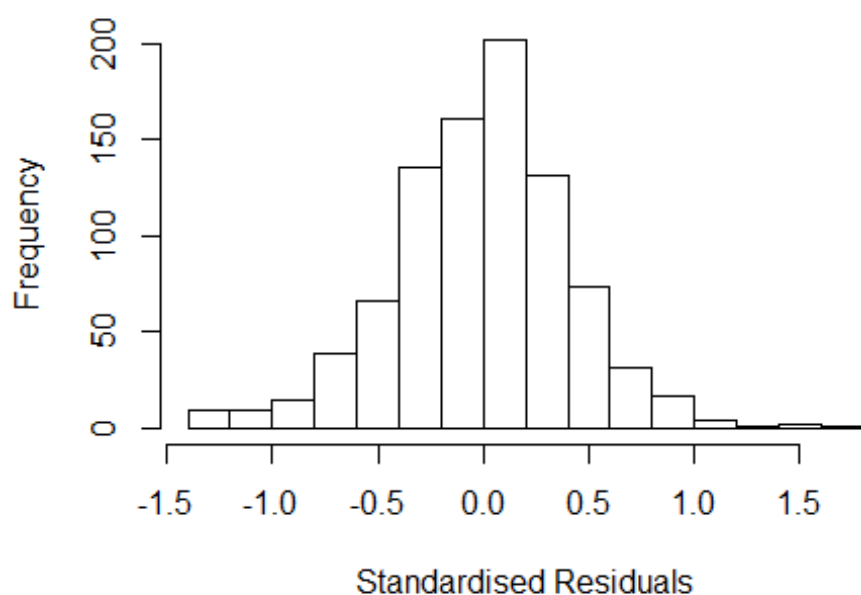
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2718 -0.2607 0.0122 0.2445 1.7062
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0704 0.1021 10.48 < 2e-16 ***
## FirstAuthorFemale1 0.0590 0.0287 2.05 0.04023 *
## LastAuthorFemale1 -0.0123 0.0285 -0.43 0.66604
## UniqueAuthors2 0.2154 0.0913 2.36 0.01845 *
## UniqueAuthors3 0.3349 0.0865 3.87 0.00012 ***
## UniqueAuthors4 0.3255 0.0904 3.60 0.00033 ***
## UniqueAuthors5 0.3433 0.0848 4.05 5.6e-05 ***
## Year1997 -0.1484 0.1216 -1.22 0.22246
## Year1998 -0.1569 0.1029 -1.52 0.12770
## Year1999 -0.2294 0.1012 -2.27 0.02363 *
```

```

## Year2000          -0.3475      0.1491    -2.33  0.02002 *
## Year2001          -0.2485      0.1243    -2.00  0.04585 *
## Year2002          -0.1965      0.0960    -2.05  0.04093 *
## Year2003          -0.3133      0.1193    -2.63  0.00876 **
## Year2004          -0.1797      0.1025    -1.75  0.08000 .
## Year2005          -0.2112      0.0899    -2.35  0.01908 *
## Year2006          -0.1950      0.0876    -2.23  0.02619 *
## Year2007          -0.2421      0.0907    -2.67  0.00772 **
## Year2008          -0.0913      0.0916    -1.00  0.31907
## Year2009          -0.1904      0.0914    -2.08  0.03756 *
## Year2010          -0.2038      0.0894    -2.28  0.02294 *
## Year2011          -0.1926      0.0899    -2.14  0.03254 *
## Year2012          -0.1805      0.0894    -2.02  0.04392 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.0781, Adjusted R-squared:  0.055
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 822 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0045 0.8690 0.9500 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          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 1.084 1          1.041
## LastAuthorFemale 1.075 1          1.037
## Year          1.142 16          1.004

```


Residuals from first and last author



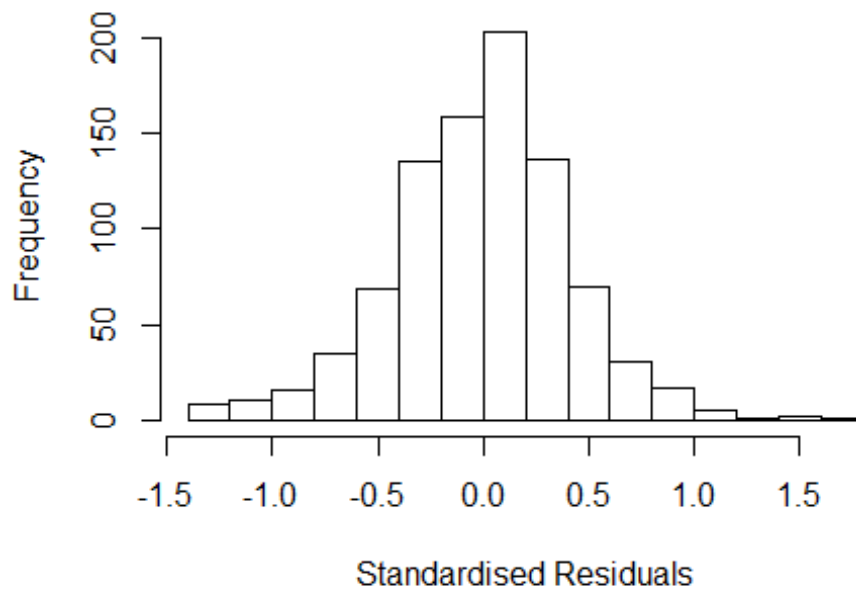
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2855 -0.2530  0.0116  0.2465  1.7494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.285178   0.083132   15.46  <2e-16 ***
## FirstAuthorFemale1  0.082174   0.028527    2.88   0.0041 **
## LastAuthorFemale1 -0.012729   0.028371   -0.45   0.6538
## Year1997        -0.111791   0.124978   -0.89   0.3713
## Year1998        -0.100245   0.107522   -0.93   0.3514
## Year1999        -0.181577   0.107458   -1.69   0.0914 .
## Year2000        -0.257332   0.163932   -1.57   0.1168
## Year2001        -0.186729   0.124851   -1.50   0.1351
## Year2002        -0.137084   0.098726   -1.39   0.1653
## Year2003        -0.232812   0.115267   -2.02   0.0437 *
## Year2004        -0.104400   0.103328   -1.01   0.3126
## Year2005        -0.133893   0.091064   -1.47   0.1418
```

```

## Year2006      -0.133877    0.090140    -1.49    0.1378
## Year2007      -0.170022    0.091942    -1.85    0.0648 .
## Year2008        0.000291    0.090919     0.00    0.9974
## Year2009      -0.095508    0.093882    -1.02    0.3093
## Year2010      -0.115173    0.090720    -1.27    0.2046
## Year2011      -0.097124    0.091027    -1.07    0.2863
## Year2012      -0.083955    0.091057    -0.92    0.3568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.0118
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 827 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8670 0.9520 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      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 1.074 1      1.036
## Year      1.074 16      1.002

```

Residuals from first author



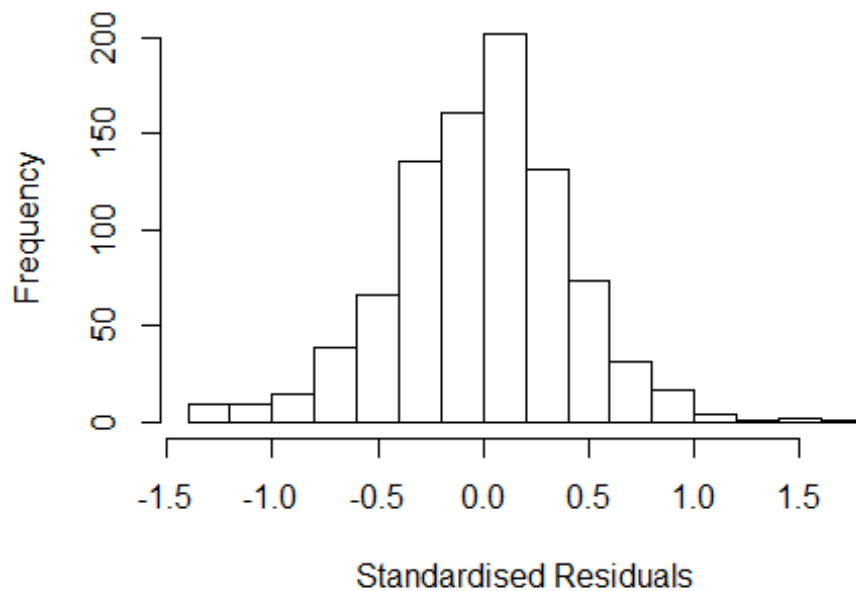
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2824 -0.2557 0.0113 0.2487 1.7389
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.282368 0.082751 15.50 <2e-16 ***
## FirstAuthorFemale1 0.080579 0.028400 2.84 0.0047 **
## Year1997 -0.110708 0.124815 -0.89 0.3753
## Year1998 -0.099222 0.107500 -0.92 0.3563
## Year1999 -0.184299 0.107359 -1.72 0.0864 .
## Year2000 -0.256748 0.163467 -1.57 0.1166
## Year2001 -0.187764 0.124993 -1.50 0.1334
## Year2002 -0.139874 0.098890 -1.41 0.1576
## Year2003 -0.232232 0.115238 -2.02 0.0442 *
## Year2004 -0.105835 0.103625 -1.02 0.3074
## Year2005 -0.133312 0.091080 -1.46 0.1436
## Year2006 -0.134880 0.090220 -1.50 0.1353
```

```

## Year2007          -0.170004    0.092089    -1.85    0.0652 .
## Year2008          -0.000779    0.091191    -0.01    0.9932
## Year2009          -0.096606    0.094131    -1.03    0.3050
## Year2010          -0.116380    0.091064    -1.28    0.2016
## Year2011          -0.099395    0.091195    -1.09    0.2761
## Year2012          -0.086413    0.091259    -0.95    0.3439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0128
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 822 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0014 0.8640 0.9520 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      1.11e-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 1.062 1          1.031
## Year            1.062 16          1.002

```

Residuals from last author



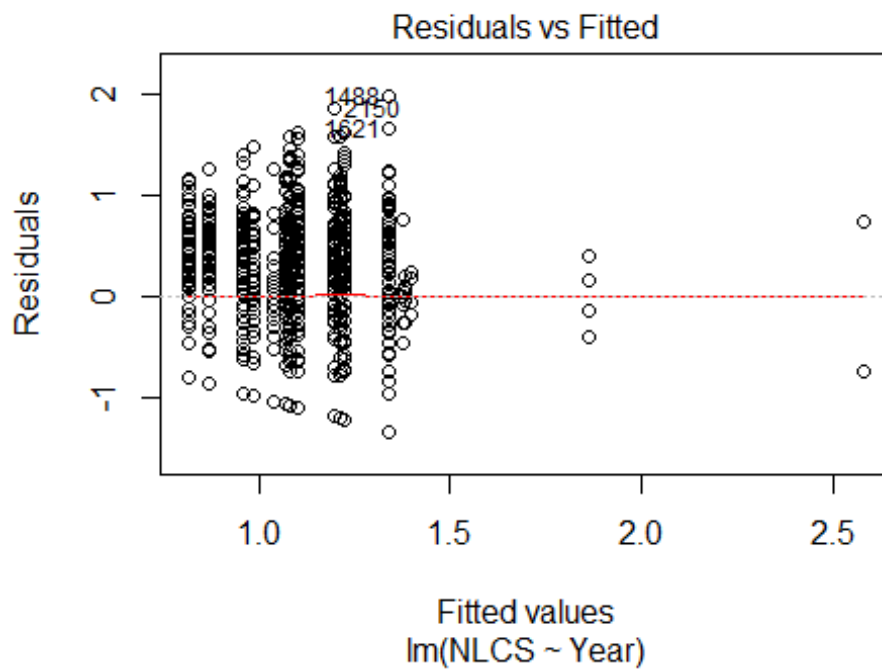
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3326 -0.2555 0.0144 0.2548 1.6828
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32411 0.07906 16.75 <2e-16 ***
## LastAuthorFemale1 -0.00226 0.02831 -0.08 0.936
## Year1997 -0.11507 0.12456 -0.92 0.356
## Year1998 -0.10541 0.10596 -0.99 0.320
## Year1999 -0.18874 0.10598 -1.78 0.075 .
## Year2000 -0.27624 0.16224 -1.70 0.089 .
## Year2001 -0.18310 0.12350 -1.48 0.139
## Year2002 -0.12128 0.09674 -1.25 0.210
## Year2003 -0.22964 0.11504 -2.00 0.046 *
## Year2004 -0.10801 0.10227 -1.06 0.291
## Year2005 -0.11662 0.08933 -1.31 0.192
## Year2006 -0.13401 0.08810 -1.52 0.129
```

```

## Year2007          -0.16620      0.09024      -1.84      0.066 .
## Year2008           0.00845      0.08910       0.09      0.924
## Year2009          -0.08615      0.09202      -0.94      0.349
## Year2010          -0.10381      0.08896      -1.17      0.244
## Year2011          -0.08932      0.08898      -1.00      0.316
## Year2012          -0.06950      0.08935      -0.78      0.437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.00344
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 835 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0125 0.8700 0.9530 0.8940 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      1.11e-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: 898"
## [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
##   69   61   52   70   65   75  104   62  160  151  153  167  188  166  153
## 2011 2012
##  172  190
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    2    5    4    7   22   90   55  138  139  137  151  165  147  134
## 2011 2012

```

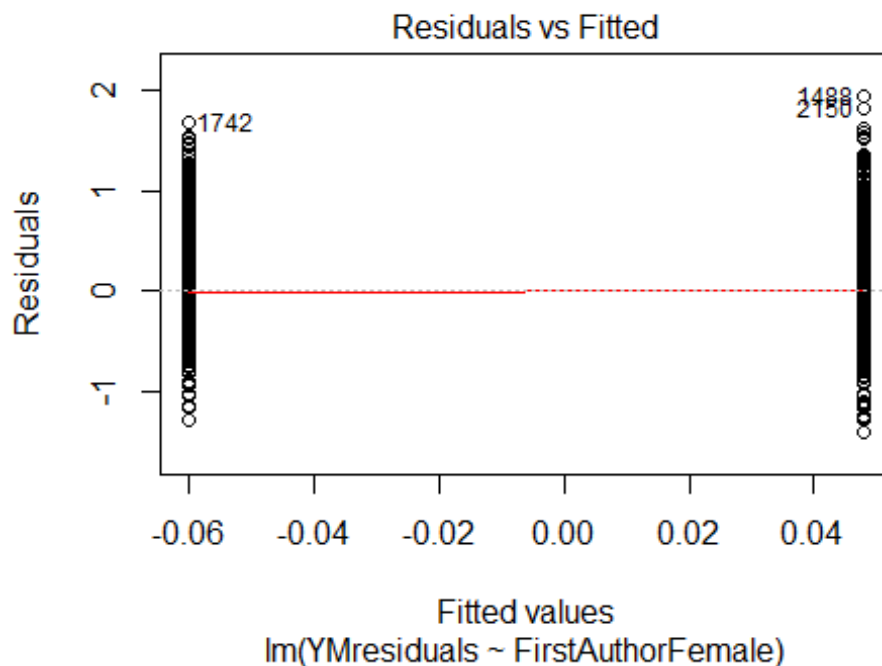
```
## 147 162
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 2 5 4 7 16 86 50 130 131 127 145 152 133 128
## 2011 2012
## 139 155
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.009
```



```
##
## 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: 3.09052066823198"
## [1] "Male first author team size 2018 geometric mean: 2.46946365955221"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

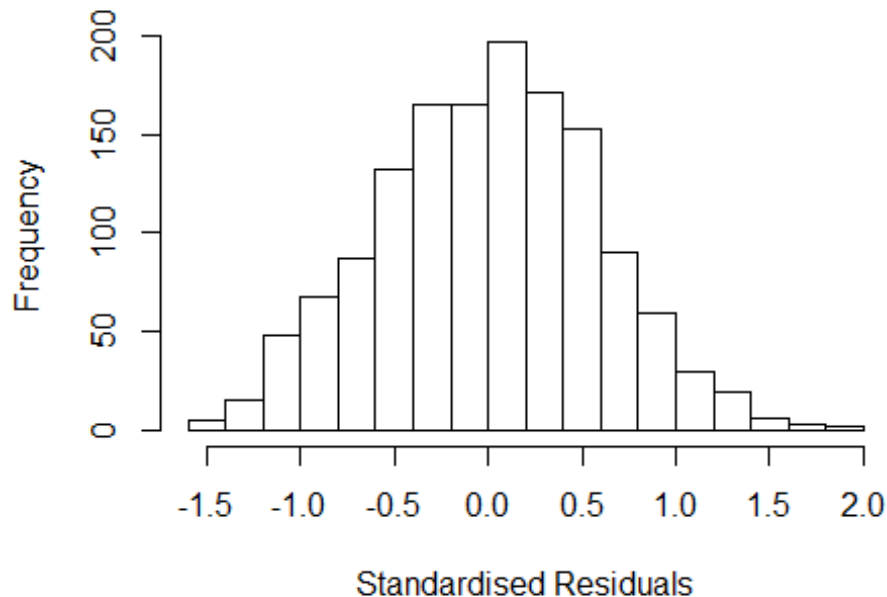
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.02890210328552"
## [1] "Male last author team size 2018 geometric mean: 2.73778970388404"

## 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 = 0.4
## 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 1.353  1      1.163
## LastAuthorFemale  1.282  1      1.132
## UniqueAuthors    1.284  4      1.032
## Year              1.298 16      1.008
```


Residuals from first and last author and team size



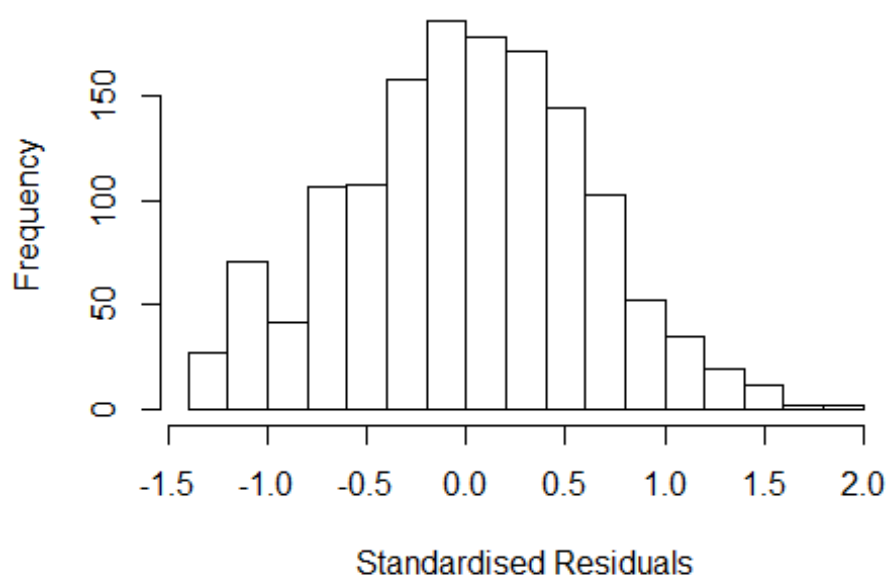
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5298 -0.4007 0.0146 0.4165 1.8269
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08166 0.16331 6.62 5.0e-11 ***
## FirstAuthorFemale1 0.04199 0.03753 1.12 0.2634
## LastAuthorFemale1 0.01140 0.03659 0.31 0.7554
## UniqueAuthors2 0.23870 0.04685 5.10 4.0e-07 ***
## UniqueAuthors3 0.19915 0.04697 4.24 2.4e-05 ***
## UniqueAuthors4 0.42414 0.05499 7.71 2.3e-14 ***
## UniqueAuthors5 0.51539 0.05488 9.39 < 2e-16 ***
## Year1997 1.25200 0.78308 1.60 0.1101
## Year1998 -0.08389 0.18626 -0.45 0.6525
## Year1999 0.36024 0.23409 1.54 0.1241
```

```

## Year2000      -0.10078    0.21161   -0.48    0.6340
## Year2001      -0.33669    0.19878   -1.69    0.0905 .
## Year2002      -0.49434    0.16779   -2.95    0.0033 **
## Year2003      -0.41839    0.18711   -2.24    0.0255 *
## Year2004      -0.43423    0.16681   -2.60    0.0093 **
## Year2005      -0.23271    0.16712   -1.39    0.1640
## Year2006      -0.35000    0.16875   -2.07    0.0382 *
## Year2007      -0.24389    0.16849   -1.45    0.1480
## Year2008      -0.23398    0.16899   -1.38    0.1664
## Year2009       0.00218    0.16905    0.01    0.9897
## Year2010      -0.10927    0.16937   -0.65    0.5189
## Year2011      -0.10333    0.16758   -0.62    0.5376
## Year2012      -0.14976    0.16742   -0.89    0.3712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.148, Adjusted R-squared:  0.134
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1294 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.327  0.875   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          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 1.285 1          1.134
## LastAuthorFemale  1.243 1          1.115
## Year              1.099 16          1.003

```

Residuals from first and last author



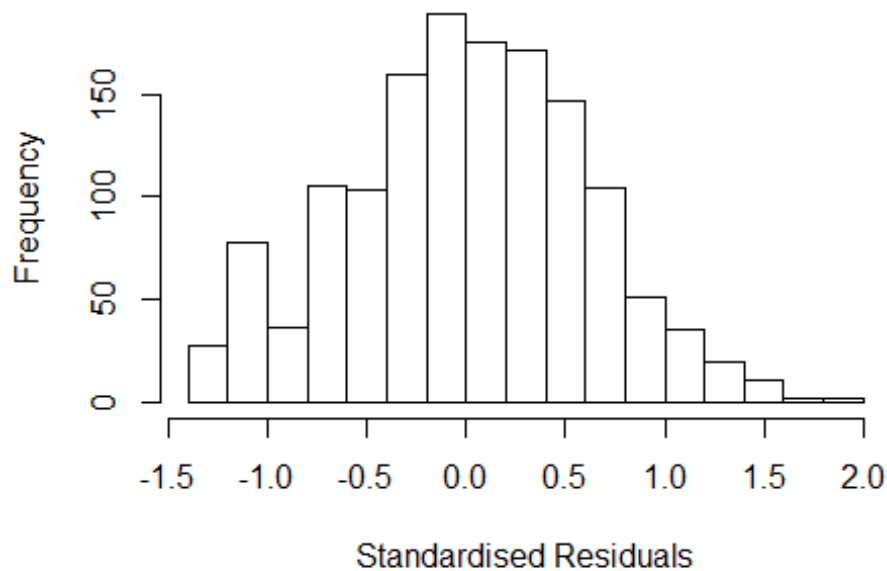
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3589 -0.3979 0.0159 0.4128 1.9671
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2745 0.0799 15.94 < 2e-16 ***
## FirstAuthorFemale1 0.1131 0.0377 3.00 0.0028 **
## LastAuthorFemale1 -0.0100 0.0375 -0.27 0.7894
## Year1997 1.1979 0.7276 1.65 0.0999 .
## Year1998 0.0869 0.1288 0.67 0.4999
## Year1999 0.5098 0.1803 2.83 0.0048 **
## Year2000 0.0764 0.1501 0.51 0.6110
## Year2001 -0.2991 0.1419 -2.11 0.0352 *
## Year2002 -0.5319 0.0969 -5.49 4.9e-08 ***
## Year2003 -0.3280 0.1199 -2.74 0.0063 **
## Year2004 -0.4751 0.0893 -5.32 1.2e-07 ***
## Year2005 -0.2525 0.0878 -2.88 0.0041 **
```

```

## Year2006          -0.3794      0.0922    -4.12  4.1e-05 ***
## Year2007          -0.2490      0.0920    -2.71  0.0069 **
## Year2008          -0.2464      0.0942    -2.61  0.0090 **
## Year2009          -0.0187      0.0904    -0.21  0.8364
## Year2010          -0.1041      0.0953    -1.09  0.2747
## Year2011          -0.1121      0.0909    -1.23  0.2180
## Year2012          -0.1264      0.0903    -1.40  0.1617
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0758, Adjusted R-squared:  0.0638
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1311 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.286  0.872  0.952  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      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 1.069 1      1.034
## Year              1.069 16      1.002

```

Residuals from first author



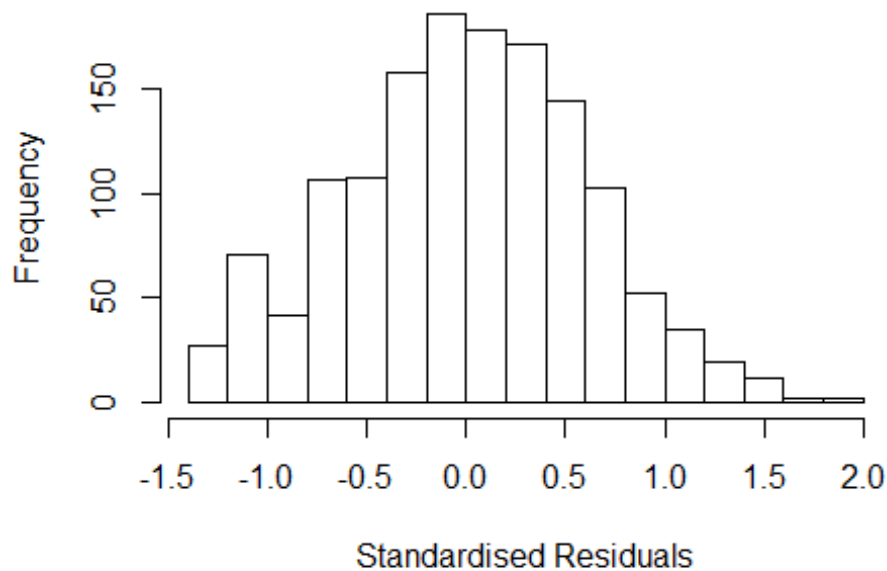
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.362 -0.396 0.017 0.411 1.964
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2728 0.0787 16.18 < 2e-16 ***
## FirstAuthorFemale1 0.1088 0.0344 3.16 0.0016 **
## Year1997 1.1989 0.7327 1.64 0.1020
## Year1998 0.0843 0.1254 0.67 0.5015
## Year1999 0.5098 0.1801 2.83 0.0047 **
## Year2000 0.0760 0.1504 0.51 0.6132
## Year2001 -0.3010 0.1413 -2.13 0.0333 *
## Year2002 -0.5320 0.0962 -5.53 3.8e-08 ***
## Year2003 -0.3278 0.1192 -2.75 0.0060 **
## Year2004 -0.4745 0.0884 -5.37 9.3e-08 ***
## Year2005 -0.2528 0.0870 -2.90 0.0037 **
## Year2006 -0.3801 0.0913 -4.16 3.4e-05 ***
```

```

## Year2007          -0.2491      0.0912   -2.73   0.0064 **
## Year2008          -0.2469      0.0935   -2.64   0.0084 **
## Year2009          -0.0193      0.0897   -0.22   0.8297
## Year2010          -0.1046      0.0945   -1.11   0.2688
## Year2011          -0.1130      0.0902   -1.25   0.2103
## Year2012          -0.1263      0.0895   -1.41   0.1583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.62
## Multiple R-squared:  0.0755, Adjusted R-squared:  0.0643
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1307 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.294  0.874  0.952  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      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 1.035 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
## -1.3374 -0.4274 0.0255 0.4294 1.9886
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.35880 0.07091 19.16 < 2e-16 ***
## LastAuthorFemale1 0.03764 0.03422 1.10 0.27161
## Year1997 1.20288 0.77191 1.56 0.11939
## Year1998 0.01906 0.09929 0.19 0.84782
## Year1999 0.48552 0.17279 2.81 0.00502 **
## Year2000 -0.00405 0.14914 -0.03 0.97833
## Year2001 -0.35364 0.13850 -2.55 0.01077 *
## Year2002 -0.57983 0.09378 -6.18 8.2e-10 ***
## Year2003 -0.36764 0.11613 -3.17 0.00158 **
## Year2004 -0.51348 0.08514 -6.03 2.1e-09 ***
## Year2005 -0.29728 0.08336 -3.57 0.00037 ***
## Year2006 -0.42755 0.08806 -4.85 1.3e-06 ***
```

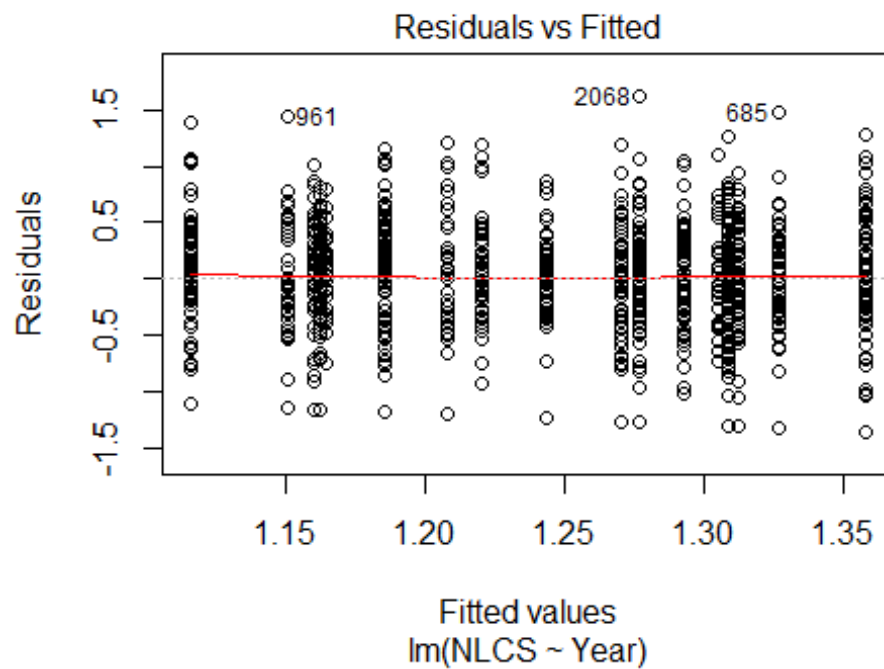
```

## Year2007          -0.28827      0.08733      -3.30   0.00099 ***
## Year2008          -0.28599      0.09036      -3.16   0.00158 **
## Year2009          -0.05903      0.08656      -0.68   0.49534
## Year2010          -0.14573      0.09298      -1.57   0.11726
## Year2011          -0.15386      0.08714      -1.77   0.07765 .
## Year2012          -0.16577      0.08634      -1.92   0.05506 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0694, Adjusted R-squared:  0.0581
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1305 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.275  0.871  0.948  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      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 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
## 112  95 128  90 123 123 104  73  94 100  81 129 134 134 116
## 2011 2012
## 111 115
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33  38  52  32  60  58  62  40  46  51  54  90  73  78  76
## 2011 2012

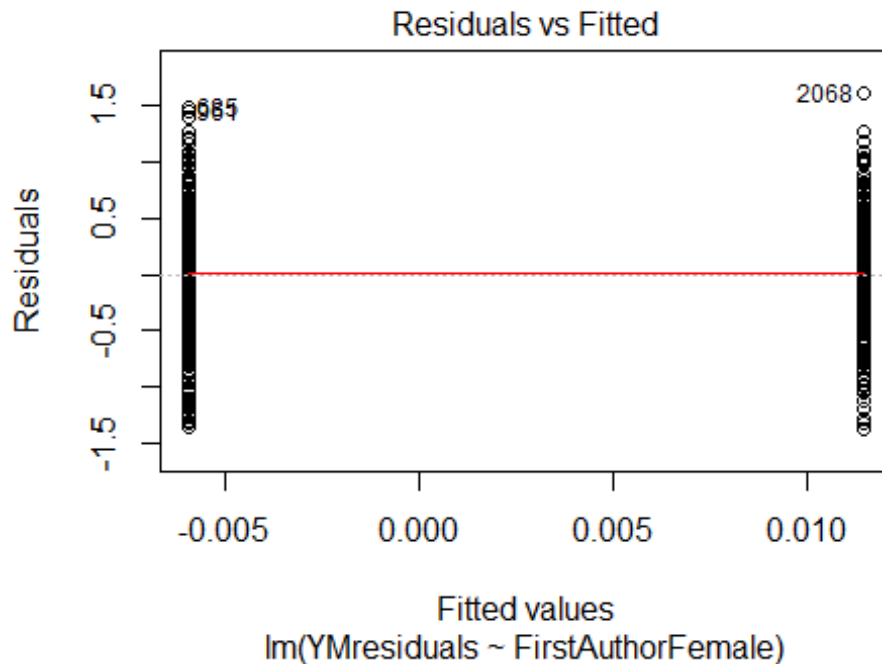
```



```
##      82      79
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   37   46   27   53   48   58   37   44   46   47   75   64   67   73
## 2011 2012
##   68   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 = 20, df = 16, p-value = 0.2
```

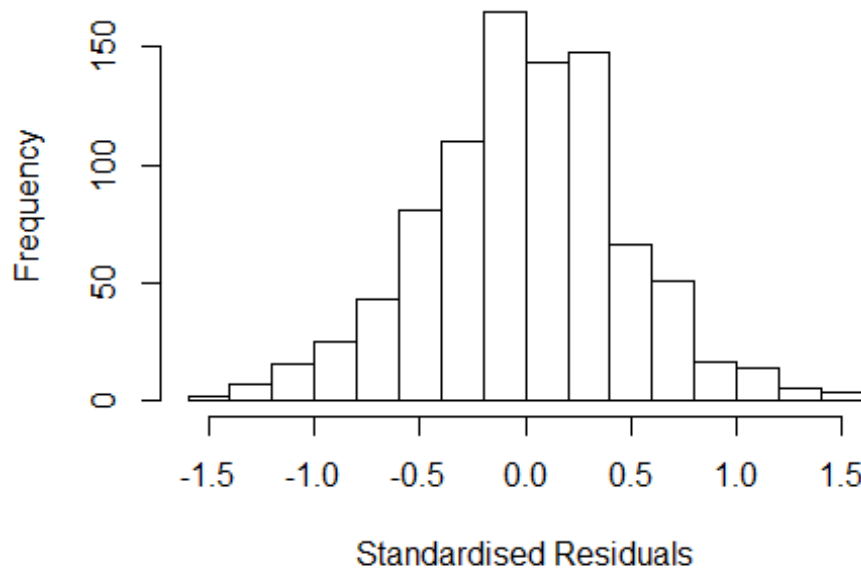


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.71673503072957"
## [1] "Male first author team size 2018 geometric mean: 4.16093996941845"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 780, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.5943403755385"
## [1] "Male last author team size 2018 geometric mean: 4.15591056045566"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 650, 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 1.096 1      1.047
## LastAuthorFemale  1.081 1      1.040
## UniqueAuthors    1.544 4      1.056
## Year              1.690 16     1.017
```

Residuals from first and last author and team size



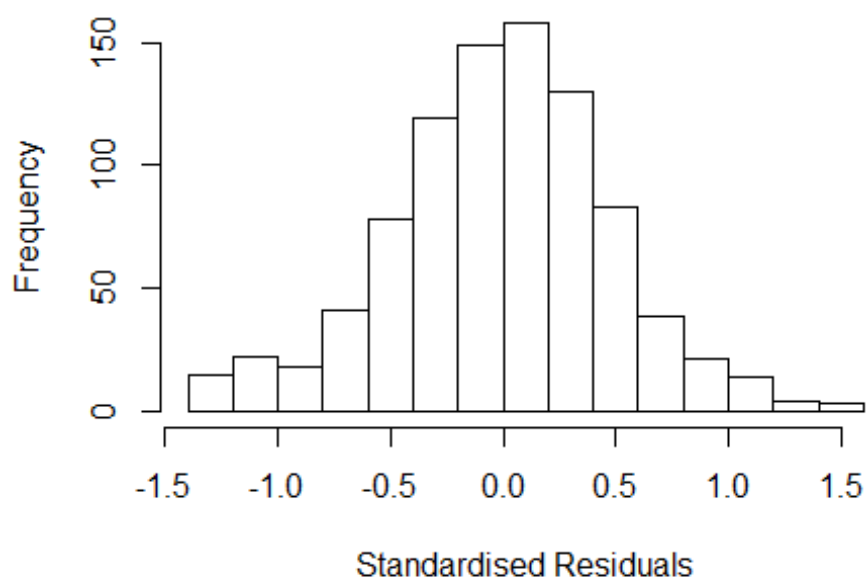
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.445319 -0.303390 -0.000817 0.303682 1.538669
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97081 0.12129 8.00 3.8e-15 ***
## FirstAuthorFemale1 -0.00585 0.03331 -0.18 0.8607
## LastAuthorFemale1 0.04405 0.04116 1.07 0.2847
## UniqueAuthors2 0.24872 0.10326 2.41 0.0162 *
## UniqueAuthors3 0.29311 0.09399 3.12 0.0019 **
## UniqueAuthors4 0.30524 0.09299 3.28 0.0011 **
## UniqueAuthors5 0.42360 0.08744 4.84 1.5e-06 ***
## Year1997 -0.07510 0.12356 -0.61 0.5435
## Year1998 -0.06539 0.10858 -0.60 0.5472
## Year1999 -0.10905 0.11004 -0.99 0.3220
```

```

## Year2000      -0.07441    0.10465   -0.71    0.4772
## Year2001      0.05091    0.09847    0.52    0.6053
## Year2002      0.03145    0.10254    0.31    0.7591
## Year2003     -0.12914    0.11254   -1.15    0.2515
## Year2004     -0.12691    0.11305   -1.12    0.2619
## Year2005     -0.05945    0.09961   -0.60    0.5508
## Year2006     -0.04917    0.10882   -0.45    0.6515
## Year2007     -0.01769    0.09757   -0.18    0.8562
## Year2008     -0.02077    0.10016   -0.21    0.8358
## Year2009     -0.09965    0.10561   -0.94    0.3456
## Year2010      0.05753    0.10297    0.56    0.5765
## Year2011     -0.03223    0.09776   -0.33    0.7417
## Year2012     -0.20502    0.10237   -2.00    0.0455 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.0764, Adjusted R-squared:  0.053
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.866  0.952  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      1.12e-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 1.081 1      1.039
## LastAuthorFemale  1.058 1      1.029
## Year              1.142 16      1.004

```

Residuals from first and last author



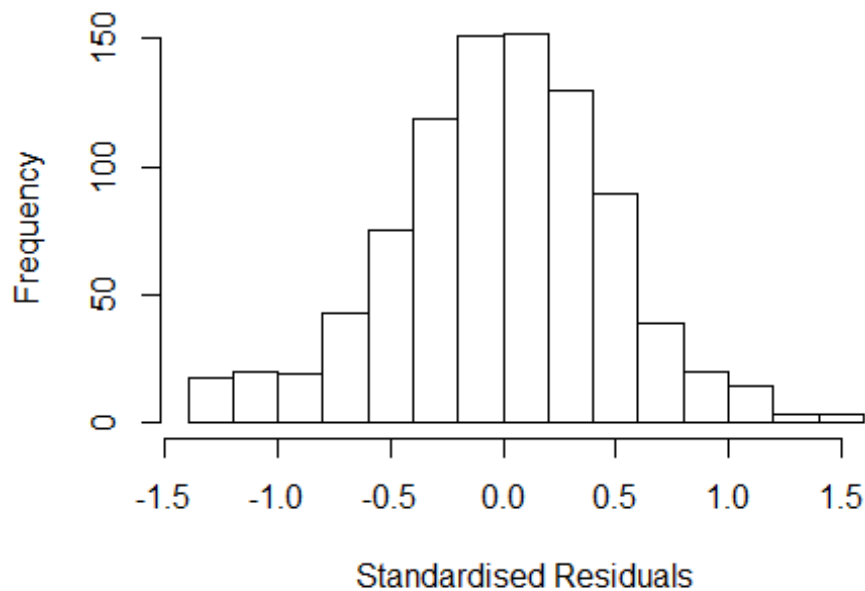
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37801 -0.30510  0.00615  0.30363  1.59956
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.30076    0.07554   17.22  <2e-16 ***
## FirstAuthorFemale1  0.01629    0.03372    0.48   0.629
## LastAuthorFemale1  0.04630    0.04177    1.11   0.268
## Year1997        -0.13924    0.12186   -1.14   0.253
## Year1998        -0.10838    0.10316   -1.05   0.294
## Year1999        -0.11465    0.10158   -1.13   0.259
## Year2000        -0.10217    0.09676   -1.06   0.291
## Year2001         0.02592    0.09793    0.26   0.791
## Year2002         0.03364    0.09932    0.34   0.735
## Year2003        -0.14462    0.11443   -1.26   0.207
## Year2004        -0.14662    0.11803   -1.24   0.215
## Year2005        -0.04437    0.09615   -0.46   0.645
```

```

## Year2006      -0.02898    0.10935   -0.27    0.791
## Year2007      0.00455    0.09369    0.05    0.961
## Year2008     -0.01473    0.09587   -0.15    0.878
## Year2009     -0.08795    0.10185   -0.86    0.388
## Year2010      0.06097    0.10184    0.60    0.550
## Year2011     -0.02160    0.09223   -0.23    0.815
## Year2012     -0.18378    0.10090   -1.82    0.069 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.00568
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.191  0.871  0.954  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      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 1.083 1      1.041
## Year              1.083 16      1.002

```

Residuals from first author



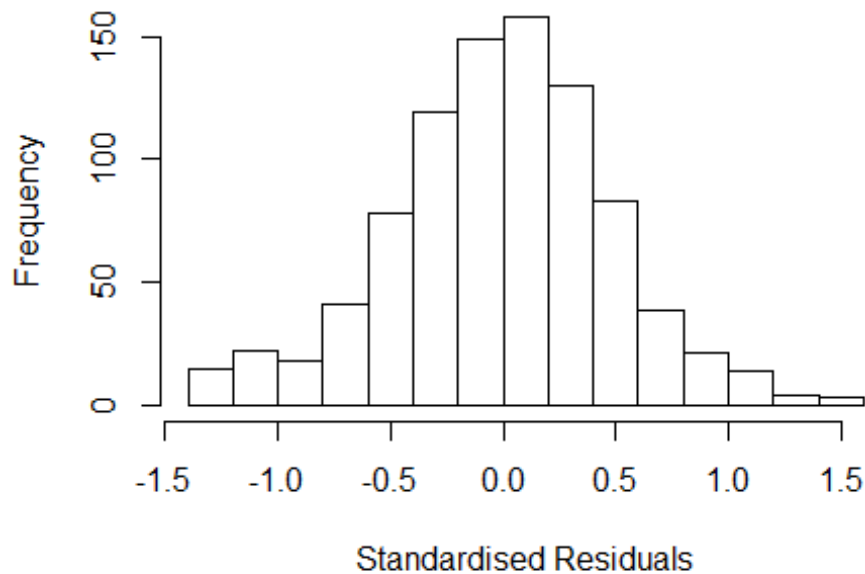
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38924 -0.30539  0.00857  0.30268  1.58885
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3093     0.0759   17.26  <2e-16 ***
## FirstAuthorFemale1  0.0196     0.0338    0.58   0.562
## Year1997         -0.1466     0.1215   -1.21   0.228
## Year1998         -0.1113     0.1030   -1.08   0.280
## Year1999         -0.1186     0.1026   -1.16   0.248
## Year2000         -0.1037     0.0972   -1.07   0.286
## Year2001          0.0239     0.0979    0.24   0.807
## Year2002          0.0302     0.0997    0.30   0.762
## Year2003         -0.1418     0.1147   -1.24   0.217
## Year2004         -0.1482     0.1182   -1.25   0.210
## Year2005         -0.0441     0.0963   -0.46   0.647
## Year2006         -0.0315     0.1095   -0.29   0.773
```

```

## Year2007          0.0031      0.0940      0.03      0.974
## Year2008          -0.0137     0.0961     -0.14     0.886
## Year2009          -0.0893     0.1020     -0.88     0.381
## Year2010           0.0603     0.1023      0.59     0.555
## Year2011          -0.0228     0.0924     -0.25     0.805
## Year2012          -0.1865     0.1017     -1.83     0.067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.0053
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 826 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.196  0.870  0.953  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      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 1.057 1      1.028
## Year            1.057 16      1.002

```


Residuals from last author



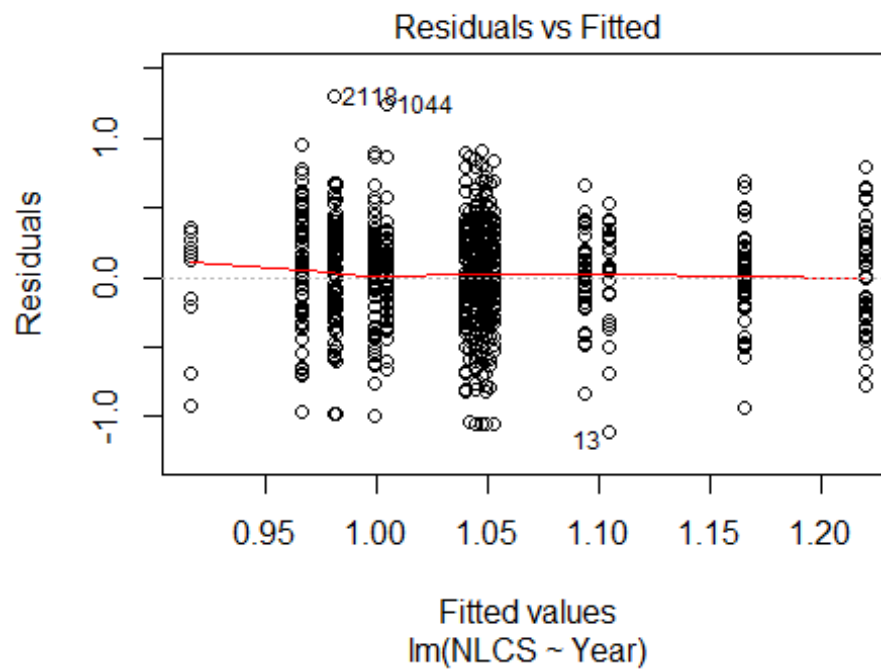
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3682 -0.3000 0.0096 0.3005 1.6111
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30479 0.07556 17.27 <2e-16 ***
## LastAuthorFemale1 0.04811 0.04184 1.15 0.251
## Year1997 -0.13749 0.12116 -1.13 0.257
## Year1998 -0.10791 0.10326 -1.04 0.296
## Year1999 -0.11332 0.10132 -1.12 0.264
## Year2000 -0.10076 0.09648 -1.04 0.297
## Year2001 0.02859 0.09710 0.29 0.768
## Year2002 0.03595 0.09910 0.36 0.717
## Year2003 -0.14276 0.11458 -1.25 0.213
## Year2004 -0.14727 0.11829 -1.24 0.213
## Year2005 -0.04133 0.09547 -0.43 0.665
## Year2006 -0.02889 0.10912 -0.26 0.791
```

```

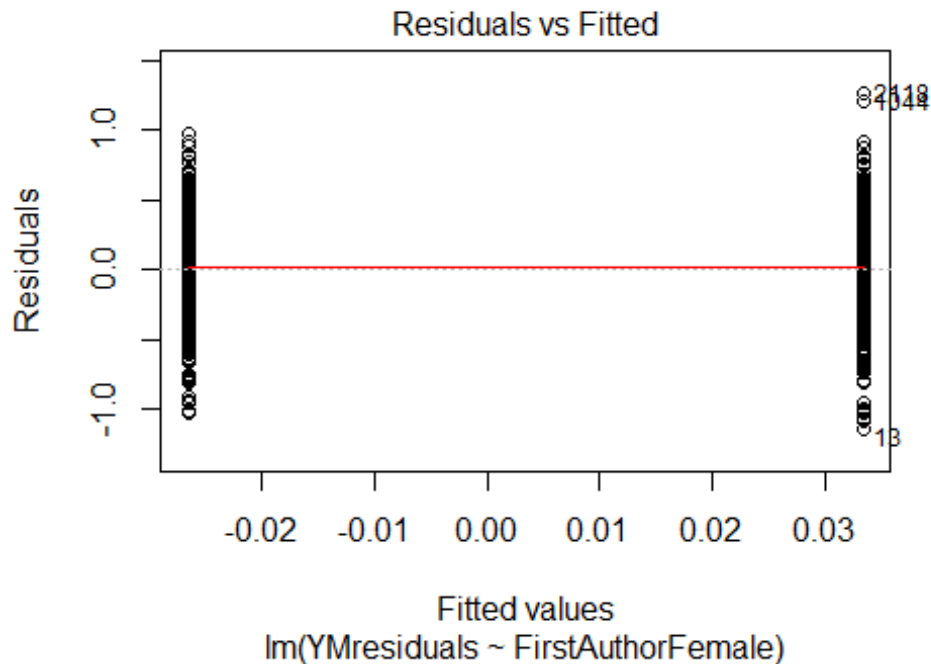
## Year2007      0.00529      0.09368      0.06      0.955
## Year2008     -0.01241      0.09580     -0.13      0.897
## Year2009     -0.08631      0.10151     -0.85      0.395
## Year2010      0.06337      0.10091      0.63      0.530
## Year2011     -0.02093      0.09209     -0.23      0.820
## Year2012     -0.18087      0.10031     -1.80      0.072 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.454
## Multiple R-squared:  0.0255, Adjusted R-squared:  0.00655
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 833 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.183  0.870  0.954  0.895  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      1.12e-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: 894"
## [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
##   59   83  106   89  105   88   94   79   97  108  117   97  119   87   88
## 2011 2012
##   87  103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   43   12   57   20   43   45   40   62   64   84   69   85   57   55
## 2011 2012

```

```
## 54 78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 42 9 53 17 38 35 31 54 52 71 62 75 52 51
## 2011 2012
## 53 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 = 27, df = 16, p-value = 0.05
```

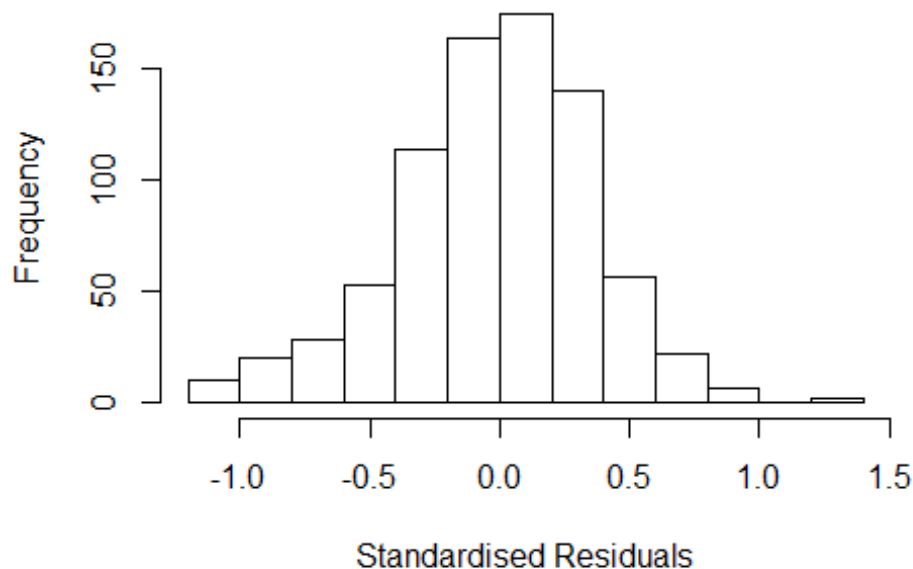


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.8, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 7.0109840249603"
## [1] "Male first author team size 2018 geometric mean: 6.47002969627394"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 820, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.71821396560233"
## [1] "Male last author team size 2018 geometric mean: 6.86037245343224"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 730, 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 1.185 1      1.088
## LastAuthorFemale  1.190 1      1.091
## UniqueAuthors    1.642 4      1.064
## Year              2.020 16     1.022
```

Residuals from first and last author and team size



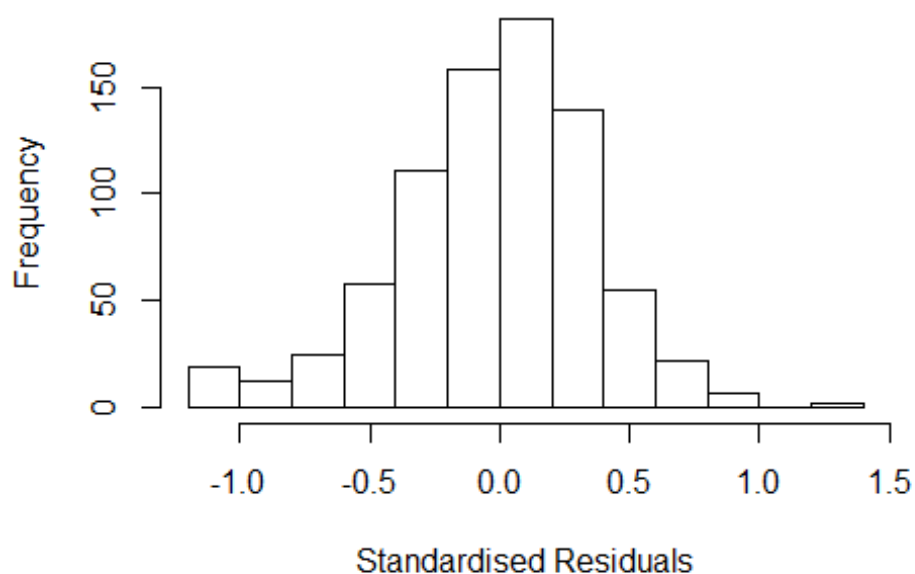
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15897 -0.24646 0.00512 0.24475 1.33924
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0327 0.1172 8.81 <2e-16 ***
## FirstAuthorFemale1 -0.0828 0.0286 -2.90 0.0039 **
## LastAuthorFemale1 -0.0412 0.0300 -1.37 0.1698
## UniqueAuthors2 0.0288 0.0847 0.34 0.7337
## UniqueAuthors3 0.1095 0.0798 1.37 0.1706
## UniqueAuthors4 0.1715 0.0819 2.09 0.0366 *
## UniqueAuthors5 0.1946 0.0762 2.55 0.0108 *
## Year1997 0.1164 0.1073 1.08 0.2785
## Year1998 -0.0960 0.1803 -0.53 0.5946
## Year1999 -0.0452 0.0999 -0.45 0.6511
```

```

## Year2000          0.0548      0.1274      0.43      0.6671
## Year2001          0.0784      0.1010      0.78      0.4375
## Year2002         -0.0134      0.0999     -0.13      0.8937
## Year2003         -0.0254      0.1001     -0.25      0.7995
## Year2004         -0.1220      0.0975     -1.25      0.2113
## Year2005         -0.0732      0.1044     -0.70      0.4832
## Year2006         -0.0536      0.0994     -0.54      0.5898
## Year2007         -0.1080      0.1004     -1.08      0.2825
## Year2008         -0.0531      0.1048     -0.51      0.6128
## Year2009         -0.0805      0.1025     -0.79      0.4324
## Year2010         -0.0388      0.1040     -0.37      0.7094
## Year2011         -0.1287      0.1047     -1.23      0.2195
## Year2012         -0.0914      0.1045     -0.87      0.3820
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0755, Adjusted R-squared:  0.049
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.874  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      1.27e-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 1.205 1      1.098
## LastAuthorFemale  1.170 1      1.082
## Year              1.382 16      1.010

```

Residuals from first and last author



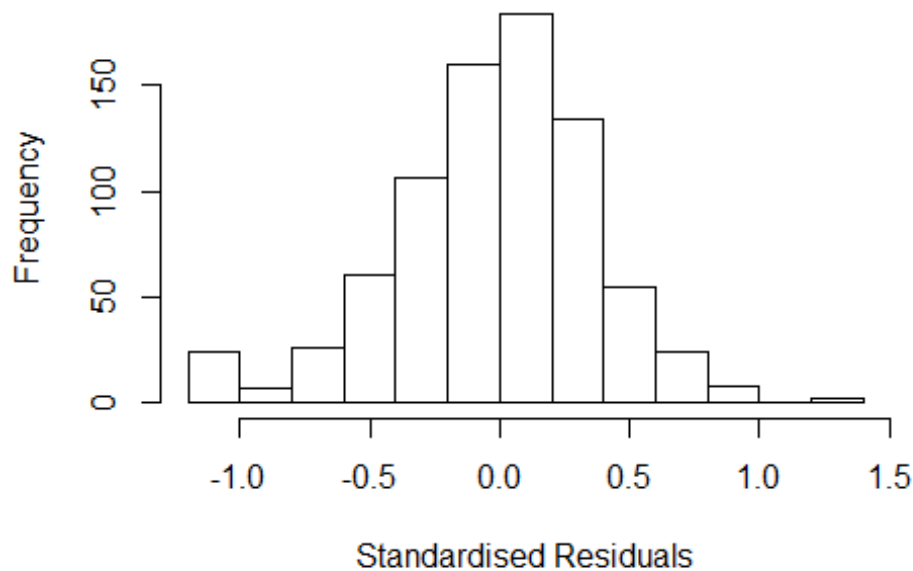
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16909 -0.24551 0.00986 0.23730 1.22747
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16909 0.08583 13.62 <2e-16 ***
## FirstAuthorFemale1 -0.07077 0.02907 -2.43 0.015 *
## LastAuthorFemale1 -0.03626 0.02990 -1.21 0.226
## Year1997 0.09748 0.10463 0.93 0.352
## Year1998 -0.13812 0.17740 -0.78 0.436
## Year1999 -0.05591 0.09680 -0.58 0.564
## Year2000 0.03499 0.13042 0.27 0.789
## Year2001 0.08020 0.09861 0.81 0.416
## Year2002 -0.00803 0.09598 -0.08 0.933
## Year2003 -0.02609 0.09813 -0.27 0.790
## Year2004 -0.13384 0.09410 -1.42 0.155
## Year2005 -0.06135 0.10035 -0.61 0.541
```

```

## Year2006      -0.05443    0.09600   -0.57    0.571
## Year2007      -0.10657    0.09754   -1.09    0.275
## Year2008      -0.04291    0.10133   -0.42    0.672
## Year2009      -0.05845    0.10032   -0.58    0.560
## Year2010      -0.03769    0.10285   -0.37    0.714
## Year2011      -0.11756    0.10245   -1.15    0.252
## Year2012      -0.07740    0.10437   -0.74    0.459
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0424, Adjusted R-squared:  0.02
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 724 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.194  0.872  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      1.27e-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 1.19 1      1.091
## Year              1.19 16      1.005

```


Residuals from first author



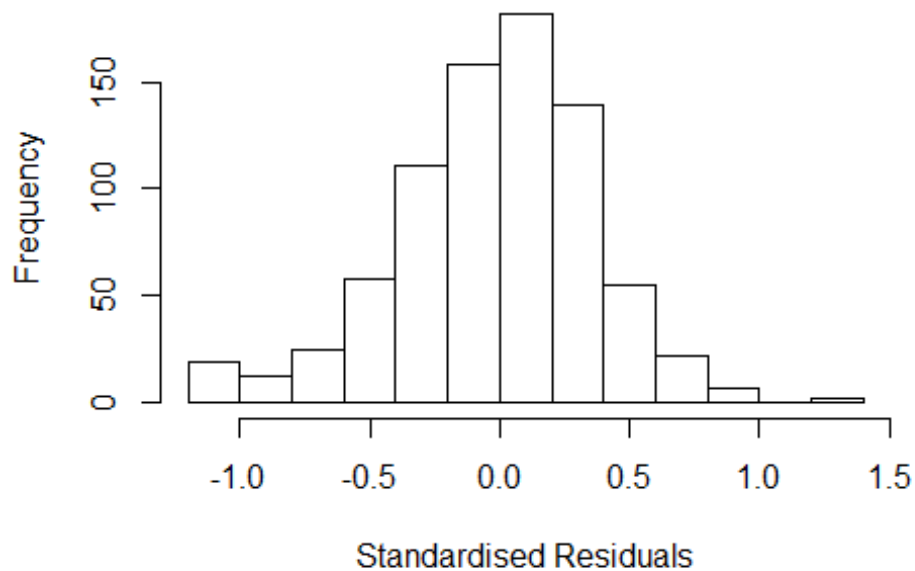
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.16761 -0.23865 0.00973 0.23215 1.23447
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1676 0.0846 13.80 <2e-16 ***
## FirstAuthorFemale1 -0.0753 0.0289 -2.61 0.0094 **
## Year1997 0.0910 0.1031 0.88 0.3780
## Year1998 -0.1520 0.1726 -0.88 0.3790
## Year1999 -0.0630 0.0956 -0.66 0.5102
## Year2000 0.0231 0.1291 0.18 0.8579
## Year2001 0.0715 0.0967 0.74 0.4596
## Year2002 -0.0100 0.0947 -0.11 0.9155
## Year2003 -0.0307 0.0971 -0.32 0.7518
## Year2004 -0.1441 0.0926 -1.56 0.1203
## Year2005 -0.0700 0.0983 -0.71 0.4769
## Year2006 -0.0643 0.0941 -0.68 0.4947
```

```

## Year2007          -0.1121      0.0967   -1.16    0.2466
## Year2008          -0.0499      0.0999   -0.50    0.6174
## Year2009          -0.0651      0.0988   -0.66    0.5101
## Year2010          -0.0488      0.1013   -0.48    0.6299
## Year2011          -0.1231      0.1011   -1.22    0.2237
## Year2012          -0.0832      0.1038   -0.80    0.4229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.04,   Adjusted R-squared:  0.0188
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 723 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.871  0.949  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      1.27e-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 1.168 1          1.081
## Year              1.168 16          1.005

```

Residuals from last author



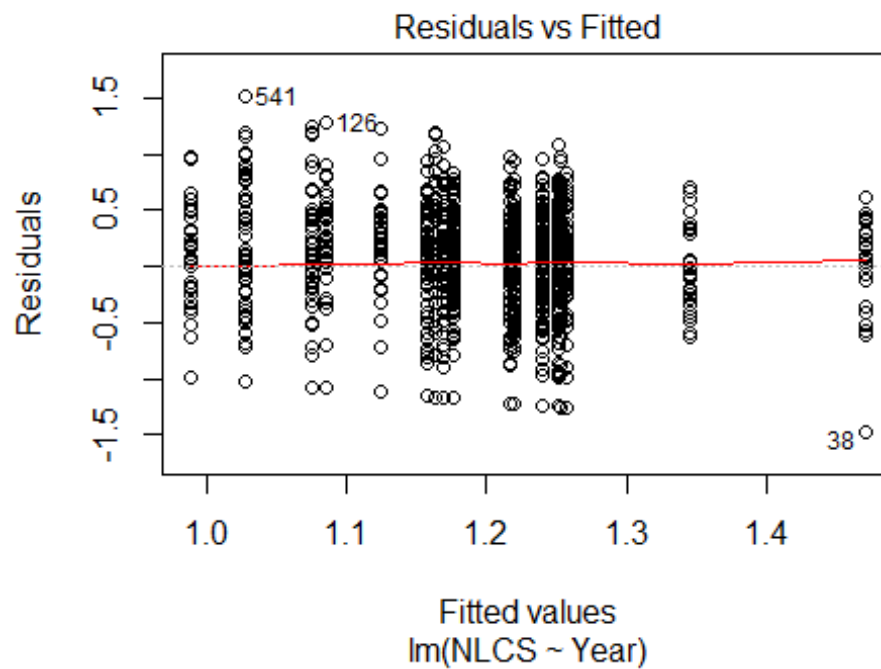
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1395 -0.2482 0.0114 0.2417 1.2732
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1395 0.0867 13.14 <2e-16 ***
## LastAuthorFemale1 -0.0463 0.0299 -1.55 0.12
## Year1997 0.1036 0.1054 0.98 0.33
## Year1998 -0.1257 0.1762 -0.71 0.48
## Year1999 -0.0569 0.0983 -0.58 0.56
## Year2000 0.0264 0.1289 0.20 0.84
## Year2001 0.0831 0.1006 0.83 0.41
## Year2002 -0.0105 0.0973 -0.11 0.91
## Year2003 -0.0230 0.0990 -0.23 0.82
## Year2004 -0.1400 0.0956 -1.47 0.14
## Year2005 -0.0760 0.1013 -0.75 0.45
## Year2006 -0.0604 0.0972 -0.62 0.53
```

```

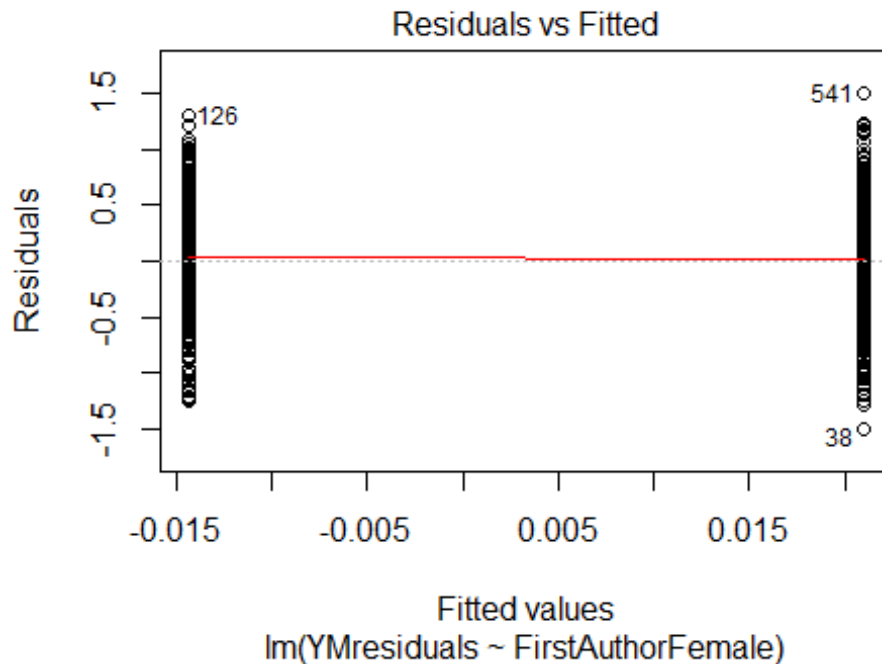
## Year2007          -0.1175      0.0982    -1.20      0.23
## Year2008          -0.0575      0.1019    -0.56      0.57
## Year2009          -0.0794      0.1012    -0.78      0.43
## Year2010          -0.0513      0.1038    -0.49      0.62
## Year2011          -0.1337      0.1037    -1.29      0.20
## Year2012          -0.0933      0.1056    -0.88      0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.352
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.0132
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.164  0.873  0.949  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      1.27e-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: 790"
## [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
##   38   37   62   62   56   86   80   86   60   81   93   98  121  140  160
## 2011 2012
##  152  177
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   26   46   48   42   57   63   74   51   69   75   87   99  115  130
## 2011 2012

```

```
## 114 135
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 22 43 43 36 53 56 69 44 66 61 80 91 98 119
## 2011 2012
## 102 115
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 52, df = 16, p-value = 1e-05
```

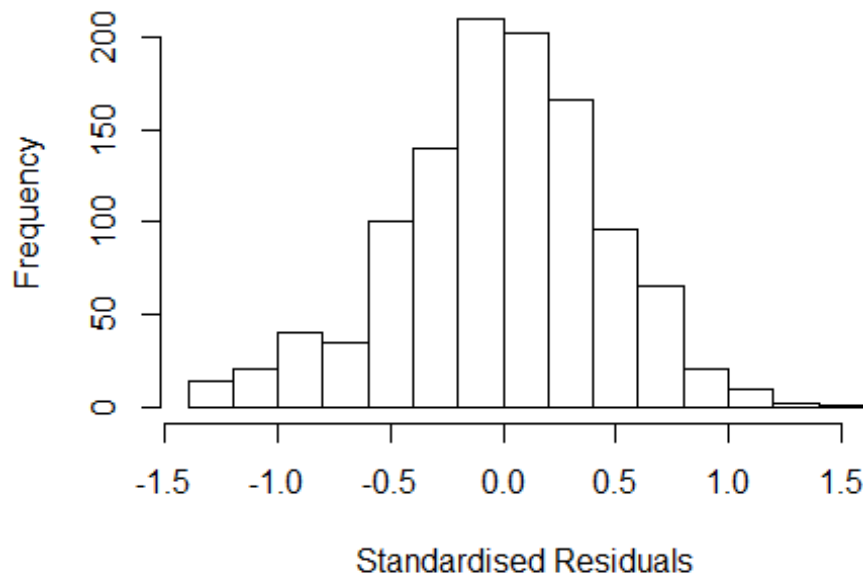


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.4, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.12608342421254"
## [1] "Male first author team size 2018 geometric mean: 5.00902985257005"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.14573886326678"
## [1] "Male last author team size 2018 geometric mean: 4.79502281006497"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, 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 1.143 1      1.069
## LastAuthorFemale  1.177 1      1.085
## UniqueAuthors    1.722 4      1.070
## Year              1.818 16     1.019
```

Residuals from first and last author and team size



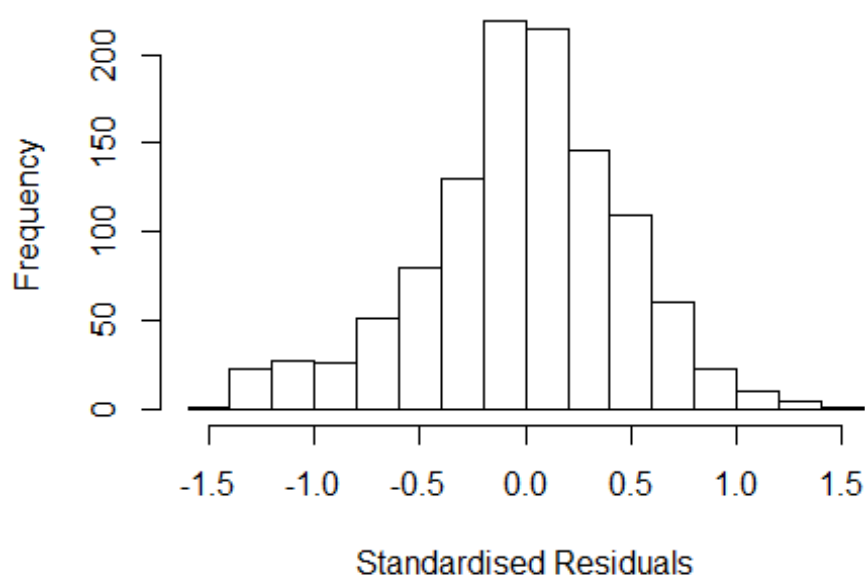
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38476 -0.28583 0.00228 0.29641 1.51021
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2508 0.0983 12.72 < 2e-16 ***
## FirstAuthorFemale1 -0.0460 0.0302 -1.52 0.12769
## LastAuthorFemale1 0.0199 0.0295 0.68 0.49920
## UniqueAuthors2 0.2335 0.0610 3.82 0.00014 ***
## UniqueAuthors3 0.2547 0.0601 4.24 2.4e-05 ***
## UniqueAuthors4 0.3075 0.0626 4.91 1.0e-06 ***
## UniqueAuthors5 0.3997 0.0581 6.88 1.0e-11 ***
## Year1997 -0.1770 0.1216 -1.46 0.14588
## Year1998 -0.3470 0.1233 -2.81 0.00498 **
## Year1999 -0.2760 0.1107 -2.49 0.01283 *
```

```

## Year2000          -0.3523      0.1263    -2.79  0.00537 **
## Year2001          -0.3469      0.1142    -3.04  0.00243 **
## Year2002          -0.2962      0.1033    -2.87  0.00422 **
## Year2003          -0.4936      0.1157    -4.27  2.1e-05 ***
## Year2004          -0.3458      0.1311    -2.64  0.00845 **
## Year2005          -0.1998      0.0962    -2.08  0.03793 *
## Year2006          -0.2549      0.1011    -2.52  0.01187 *
## Year2007          -0.2657      0.0974    -2.73  0.00647 **
## Year2008          -0.2303      0.0910    -2.53  0.01153 *
## Year2009          -0.2829      0.0883    -3.20  0.00140 **
## Year2010          -0.2789      0.0905    -3.08  0.00211 **
## Year2011          -0.3161      0.0919    -3.44  0.00061 ***
## Year2012          -0.3472      0.0900    -3.86  0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0943, Adjusted R-squared:  0.0762
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 1033 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.198  0.864  0.950  0.895  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      8.90e-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 1.114 1      1.056
## LastAuthorFemale  1.125 1      1.061
## Year              1.183 16      1.005

```


Residuals from first and last author



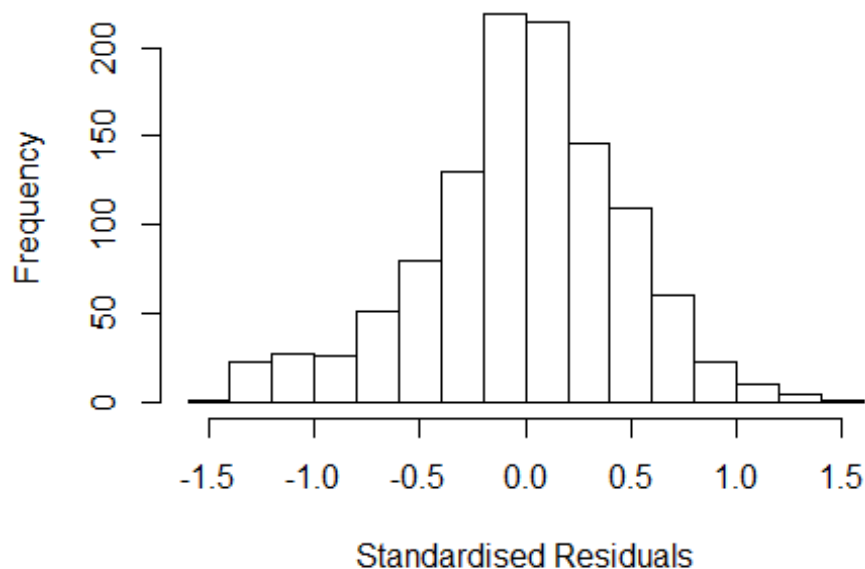
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49188 -0.28359  0.00704  0.29790  1.52820
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.491881    0.076326   19.55 < 2e-16 ***
## FirstAuthorFemale1 -0.037784    0.030900   -1.22  0.22166
## LastAuthorFemale1 -0.000749    0.029581   -0.03  0.97980
## Year1997        -0.170429    0.112894   -1.51  0.13142
## Year1998        -0.410228    0.121278   -3.38  0.00074 ***
## Year1999        -0.338538    0.103671   -3.27  0.00113 **
## Year2000        -0.440379    0.118916   -3.70  0.00022 ***
## Year2001        -0.392580    0.114293   -3.43  0.00061 ***
## Year2002        -0.273944    0.096409   -2.84  0.00457 **
## Year2003        -0.477334    0.114812   -4.16  3.5e-05 ***
## Year2004        -0.292765    0.125573   -2.33  0.01991 *
## Year2005        -0.179037    0.091726   -1.95  0.05120 .
```

```

## Year2006      -0.219162    0.097546    -2.25    0.02485 *
## Year2007      -0.241993    0.095896    -2.52    0.01176 *
## Year2008      -0.187473    0.085269    -2.20    0.02811 *
## Year2009      -0.242169    0.081669    -2.97    0.00309 **
## Year2010      -0.206464    0.083198    -2.48    0.01323 *
## Year2011      -0.232345    0.085121    -2.73    0.00644 **
## Year2012      -0.268658    0.081185    -3.31    0.00097 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.025
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1032 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.190  0.858  0.950  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      8.90e-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 1.076 1      1.037
## Year      1.076 16      1.002

```

Residuals from first author



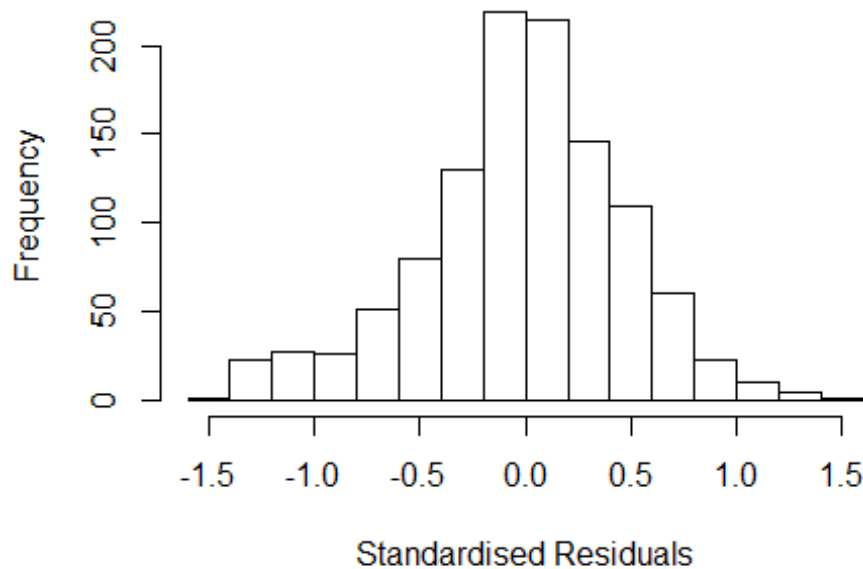
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.49159 -0.28393 0.00663 0.29777 1.52782
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4916 0.0736 20.28 < 2e-16 ***
## FirstAuthorFemale1 -0.0380 0.0304 -1.25 0.21255
## Year1997 -0.1705 0.1127 -1.51 0.13080
## Year1998 -0.4102 0.1214 -3.38 0.00075 ***
## Year1999 -0.3384 0.1032 -3.28 0.00107 **
## Year2000 -0.4403 0.1189 -3.70 0.00022 ***
## Year2001 -0.3925 0.1137 -3.45 0.00058 ***
## Year2002 -0.2738 0.0957 -2.86 0.00429 **
## Year2003 -0.4774 0.1147 -4.16 3.4e-05 ***
## Year2004 -0.2927 0.1252 -2.34 0.01955 *
## Year2005 -0.1789 0.0913 -1.96 0.05016 .
## Year2006 -0.2191 0.0972 -2.25 0.02443 *
```

```

## Year2007          -0.2418      0.0956   -2.53  0.01159 *
## Year2008          -0.1873      0.0844   -2.22  0.02672 *
## Year2009          -0.2421      0.0808   -2.99  0.00281 **
## Year2010          -0.2063      0.0823   -2.51  0.01236 *
## Year2011          -0.2321      0.0835   -2.78  0.00550 **
## Year2012          -0.2686      0.0806   -3.33  0.00090 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.0259
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1032 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.189  0.858  0.950  0.889  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      8.90e-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 1.088 1      1.043
## Year      1.088 16      1.003

```

Residuals from last author



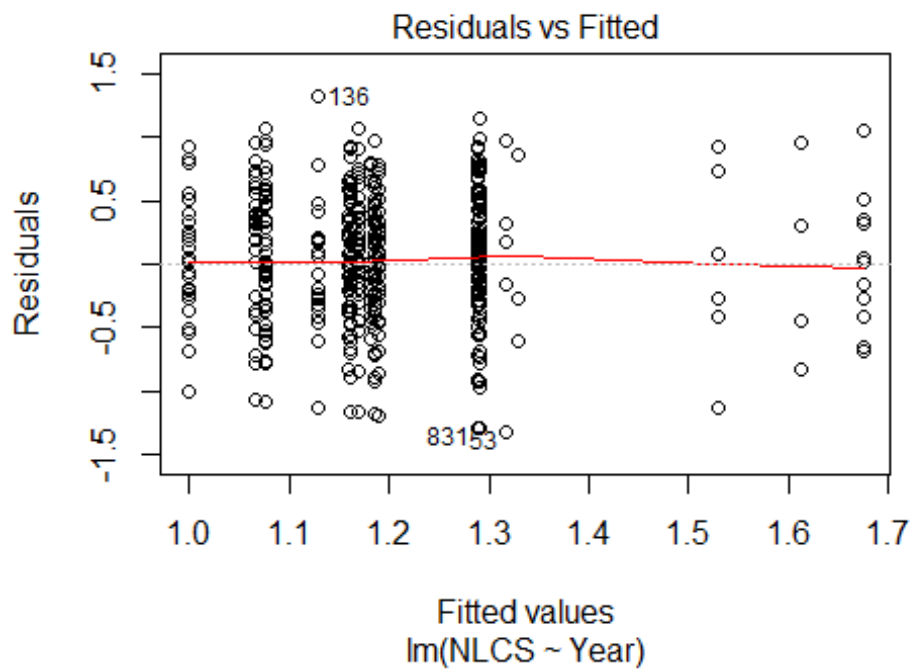
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.48043 -0.29269  0.00468  0.29676  1.55395
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.48043    0.07534   19.65 < 2e-16 ***
## LastAuthorFemale1 -0.00986    0.02920   -0.34  0.73575
## Year1997        -0.17097    0.11264   -1.52  0.12934
## Year1998        -0.41417    0.12216   -3.39  0.00072 ***
## Year1999        -0.34584    0.10237   -3.38  0.00075 ***
## Year2000        -0.43731    0.11819   -3.70  0.00023 ***
## Year2001        -0.39903    0.11464   -3.48  0.00052 ***
## Year2002        -0.27753    0.09631   -2.88  0.00403 **
## Year2003        -0.48252    0.11463   -4.21  2.8e-05 ***
## Year2004        -0.29932    0.12476   -2.40  0.01660 *
## Year2005        -0.18645    0.09133   -2.04  0.04145 *
## Year2006        -0.22790    0.09645   -2.36  0.01830 *
```

```

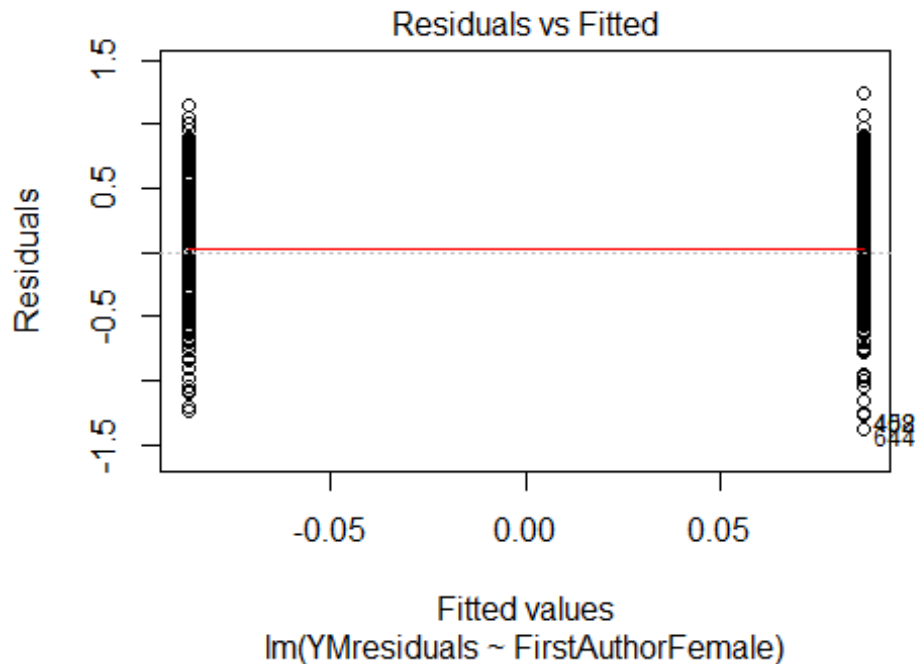
## Year2007          -0.25259      0.09507    -2.66  0.00800 **
## Year2008          -0.19429      0.08473    -2.29  0.02203 *
## Year2009          -0.25045      0.08112    -3.09  0.00207 **
## Year2010          -0.21560      0.08257    -2.61  0.00915 **
## Year2011          -0.24416      0.08408    -2.90  0.00376 **
## Year2012          -0.27765      0.08037    -3.45  0.00057 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.0391, Adjusted R-squared:  0.0243
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 93 weights are ~= 1. The remaining 1031 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.174  0.858  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      8.90e-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: 1124"
## [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
##    8   13   12   18   22   34   28   39   22   31   49   57   59   61   66
## 2011 2012
##   73  109
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    4    7    5   13   19   24   31   14   27   32   45   53   51   50
## 2011 2012

```

```
## 55 79
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 4 7 5 12 18 23 28 12 24 29 42 49 46 41
## 2011 2012
## 50 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 = 18, df = 16, p-value = 0.3
```

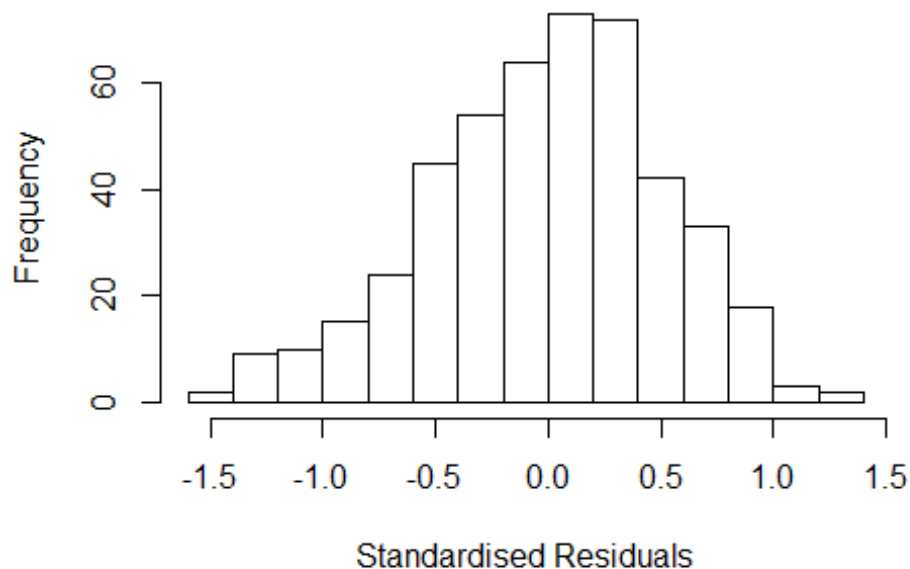


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.9930395148147"
## [1] "Male first author team size 2018 geometric mean: 3.80110901973007"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.93815282844642"
## [1] "Male last author team size 2018 geometric mean: 4.15552046097582"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, 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 1.550 1          1.245
## LastAuthorFemale  1.477 1          1.215
## UniqueAuthors     2.092 4          1.097
## Year               2.720 16         1.032
```


Residuals from first and last author and team size



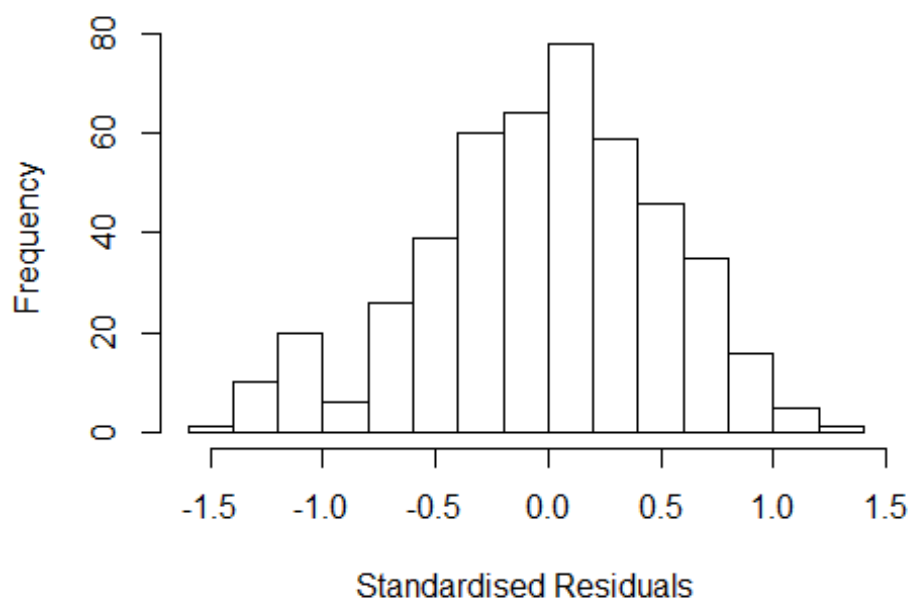
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4022 -0.3602 0.0202 0.3453 1.2443
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1527 0.4033 2.86 0.00447 **
## FirstAuthorFemale1 0.1612 0.0591 2.73 0.00664 **
## LastAuthorFemale1 0.0174 0.0576 0.30 0.76298
## UniqueAuthors2 0.1680 0.0972 1.73 0.08471 .
## UniqueAuthors3 0.2204 0.0912 2.42 0.01599 *
## UniqueAuthors4 0.2621 0.0977 2.68 0.00760 **
## UniqueAuthors5 0.3283 0.0850 3.86 0.00013 ***
## Year1997 0.3314 0.6936 0.48 0.63307
## Year1998 0.2482 0.4913 0.51 0.61369
## Year1999 0.0812 0.5213 0.16 0.87623
```

```

## Year2000          0.3290      0.4454      0.74  0.46053
## Year2001          -0.1992      0.4191     -0.48  0.63476
## Year2002          -0.3387      0.4197     -0.81  0.42004
## Year2003          -0.2395      0.4208     -0.57  0.56954
## Year2004          -0.1937      0.4306     -0.45  0.65301
## Year2005          -0.3644      0.4225     -0.86  0.38887
## Year2006          -0.2379      0.4317     -0.55  0.58190
## Year2007          -0.2400      0.4211     -0.57  0.56905
## Year2008          -0.2287      0.4172     -0.55  0.58380
## Year2009          -0.4143      0.4194     -0.99  0.32373
## Year2010          -0.3238      0.4206     -0.77  0.44182
## Year2011          -0.1532      0.4225     -0.36  0.71702
## Year2012          -0.2248      0.4186     -0.54  0.59163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.117, Adjusted R-squared:  0.0728
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 431 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.443  0.875  0.953  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      2.15e-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 1.568 1      1.252
## LastAuthorFemale  1.368 1      1.170
## Year              1.422 16      1.011

```

Residuals from first and last author



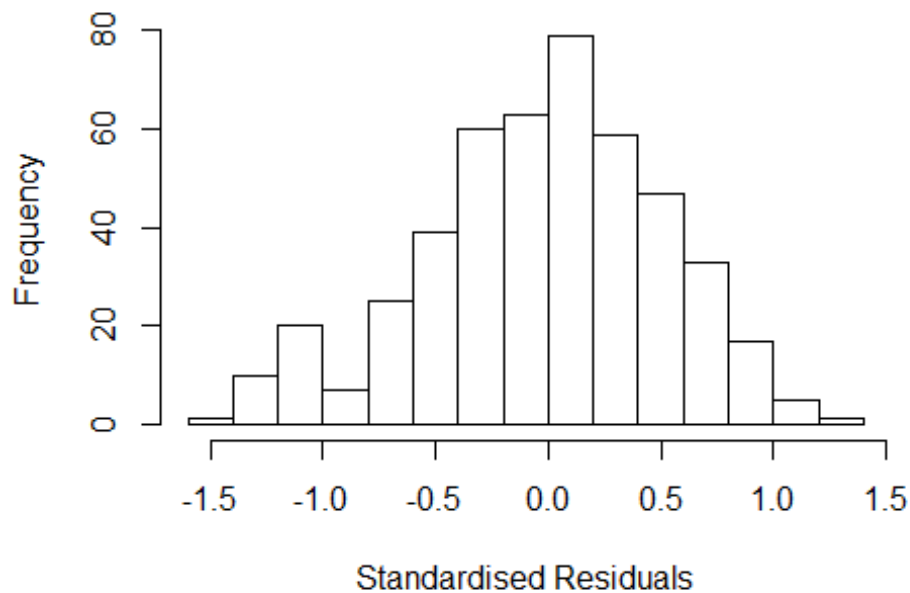
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4223 -0.3673 0.0351 0.3581 1.2603
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15755 0.47924 2.42 0.0161 *
## FirstAuthorFemale1 0.19516 0.06189 3.15 0.0017 **
## LastAuthorFemale1 0.01040 0.05798 0.18 0.8578
## Year1997 0.41973 0.67179 0.62 0.5324
## Year1998 0.32587 0.55264 0.59 0.5557
## Year1999 0.19684 0.56736 0.35 0.7288
## Year2000 0.41081 0.50468 0.81 0.4161
## Year2001 -0.06681 0.48559 -0.14 0.8906
## Year2002 -0.16936 0.48345 -0.35 0.7263
## Year2003 -0.03989 0.48266 -0.08 0.9342
## Year2004 -0.06908 0.49968 -0.14 0.8901
## Year2005 -0.19876 0.48790 -0.41 0.6839
```

```

## Year2006      -0.07329      0.49485      -0.15      0.8823
## Year2007      -0.04765      0.48346      -0.10      0.9215
## Year2008      -0.04143      0.48133      -0.09      0.9314
## Year2009      -0.22545      0.48279      -0.47      0.6407
## Year2010      -0.09887      0.48105      -0.21      0.8373
## Year2011       0.06963      0.48390       0.14      0.8857
## Year2012      -0.00642      0.48116      -0.01      0.9894
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0813, Adjusted R-squared:  0.0443
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 431 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.436  0.870   0.951   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      2.15e-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 1.264 1      1.124
## Year              1.264 16      1.007

```

Residuals from first author



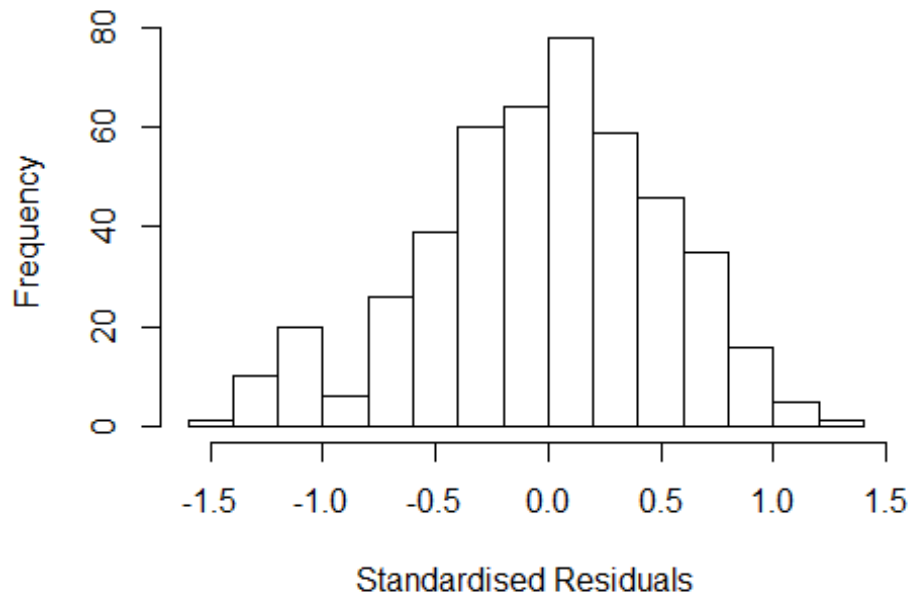
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4279 -0.3714 0.0354 0.3542 1.2634
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15929 0.48062 2.41 0.0163 *
## FirstAuthorFemale1 0.19885 0.05577 3.57 0.0004 ***
## Year1997 0.41740 0.67346 0.62 0.5357
## Year1998 0.32540 0.55324 0.59 0.5567
## Year1999 0.19491 0.56922 0.34 0.7322
## Year2000 0.40895 0.50590 0.81 0.4193
## Year2001 -0.06539 0.48811 -0.13 0.8935
## Year2002 -0.16756 0.48577 -0.34 0.7303
## Year2003 -0.03911 0.48531 -0.08 0.9358
## Year2004 -0.07018 0.50115 -0.14 0.8887
## Year2005 -0.19867 0.48995 -0.41 0.6853
## Year2006 -0.07261 0.49836 -0.15 0.8842
```

```

## Year2007          -0.04771    0.48595   -0.10    0.9218
## Year2008          -0.04081    0.48417   -0.08    0.9329
## Year2009          -0.22525    0.48512   -0.46    0.6426
## Year2010          -0.09718    0.48433   -0.20    0.8411
## Year2011           0.06976    0.48724    0.14    0.8862
## Year2012          -0.00633    0.48348   -0.01    0.9896
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.0815, Adjusted R-squared:  0.0466
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 432 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.428  0.869  0.951  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      2.15e-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 1.102 1      1.050
## Year              1.102 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
## -1.4485 -0.3390 0.0154 0.3383 1.3268
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2191 0.4621 2.64 0.0086 **
## LastAuthorFemale1 0.0839 0.0519 1.62 0.1068
## Year1997 0.3582 0.6605 0.54 0.5879
## Year1998 0.3098 0.5429 0.57 0.5685
## Year1999 0.2293 0.5797 0.40 0.6926
## Year2000 0.4355 0.4824 0.90 0.3672
## Year2001 -0.1047 0.4697 -0.22 0.8237
## Year2002 -0.1759 0.4679 -0.38 0.7072
## Year2003 -0.0618 0.4659 -0.13 0.8945
## Year2004 -0.0526 0.4841 -0.11 0.9135
## Year2005 -0.1839 0.4710 -0.39 0.6964
## Year2006 -0.0788 0.4815 -0.16 0.8701
```

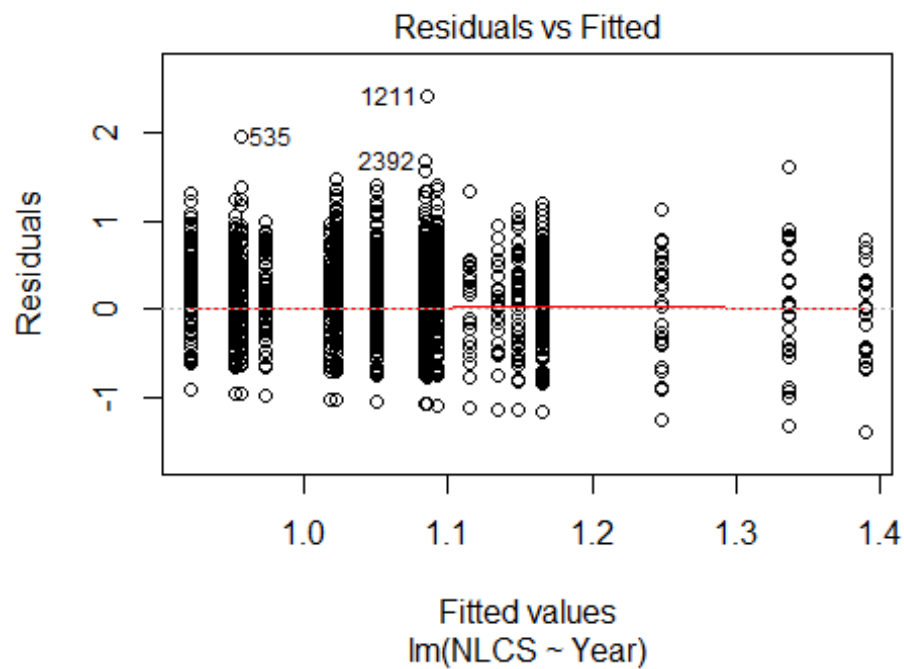
```

## Year2007          -0.0357      0.4673   -0.08    0.9392
## Year2008          -0.0419      0.4654   -0.09    0.9283
## Year2009          -0.1930      0.4670   -0.41    0.6796
## Year2010          -0.0889      0.4659   -0.19    0.8488
## Year2011           0.0955      0.4690    0.20    0.8387
## Year2012           0.0476      0.4647    0.10    0.9185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0584, Adjusted R-squared:  0.0227
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 427 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.420  0.860   0.949   0.901   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      2.15e-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: 466"
## [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
##   40   44   50   49   71  114  124  108  109  193  188  209  314  400  363
## 2011 2012
##  381  351
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21   25   22   29   31   52   99   88   83  149  156  185  267  343  300
## 2011 2012

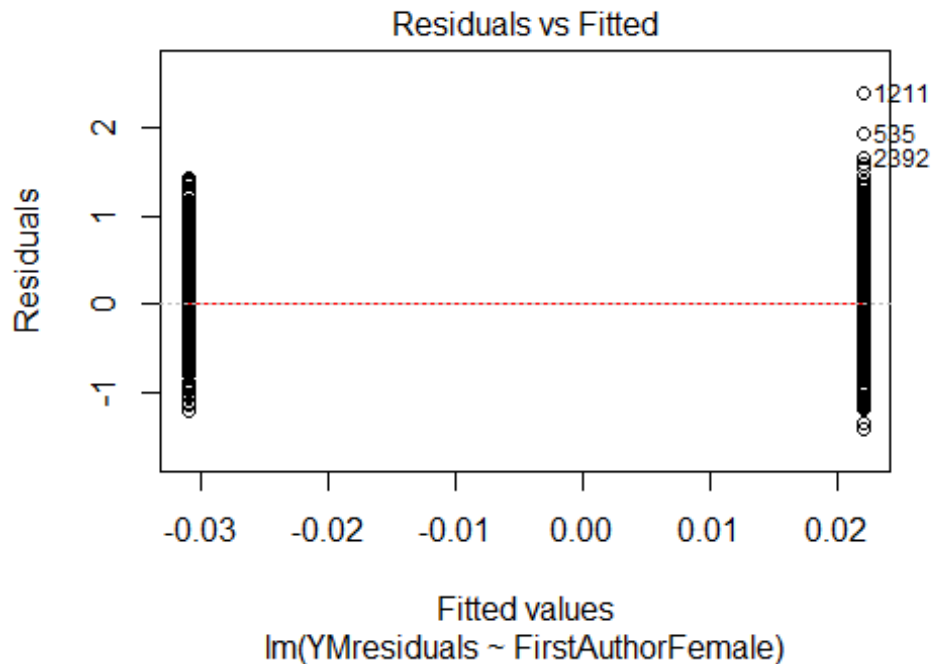
```



```
## 333 291
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 22 21 24 28 46 90 80 74 134 138 176 243 319 276
## 2011 2012
## 311 267
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 13, df = 16, p-value = 0.7
```

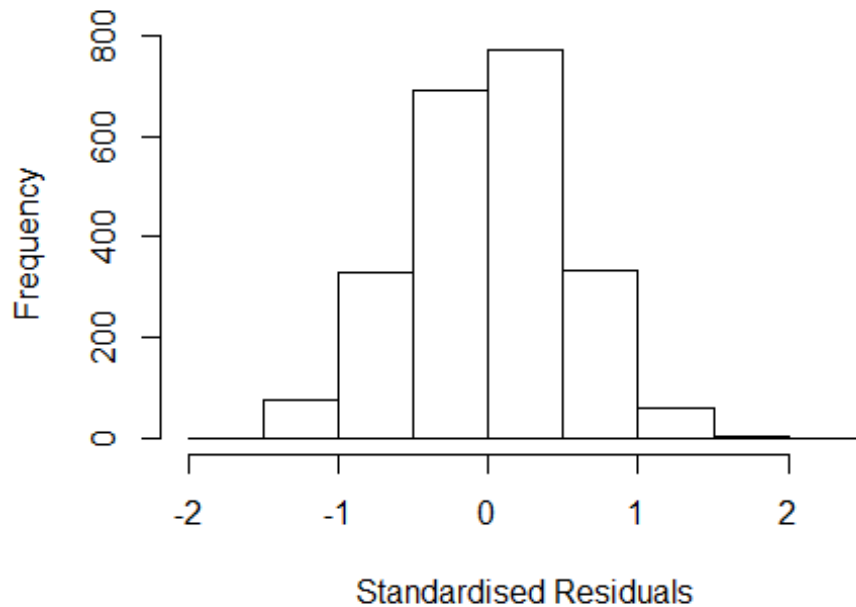


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.9, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.98747923907292"
## [1] "Male first author team size 2018 geometric mean: 3.17308445473381"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.81900106393294"
## [1] "Male last author team size 2018 geometric mean: 3.56138427767732"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, 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 1.261 1      1.123
## LastAuthorFemale  1.203 1      1.097
## UniqueAuthors     1.221 4      1.025
## Year              1.253 16      1.007
```

Residuals from first and last author and team size



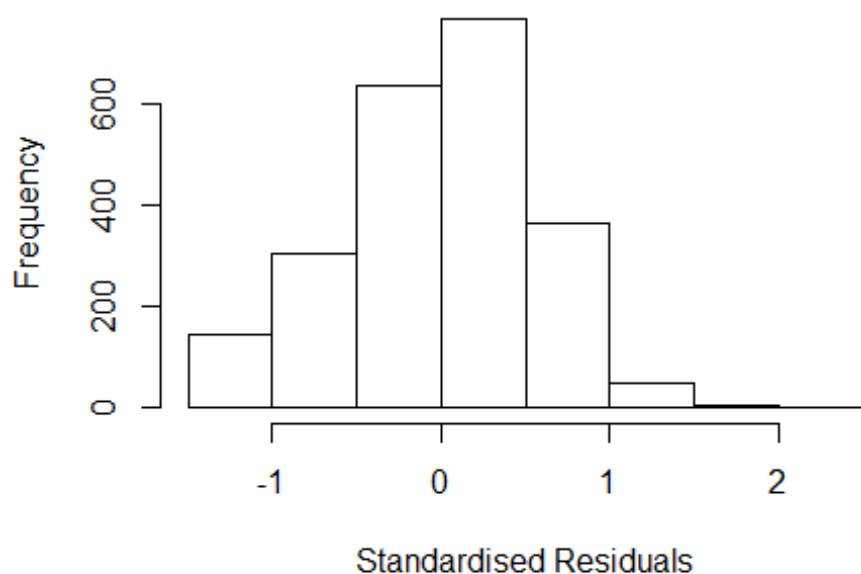
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5661 -0.3767  0.0234  0.3570  2.3737
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2723    0.1333   9.54 < 2e-16 ***
## FirstAuthorFemale1  0.0136    0.0266   0.51  0.61068
## LastAuthorFemale1 -0.0108    0.0253  -0.43  0.66866
## UniqueAuthors2     0.2802    0.0384   7.30 3.8e-13 ***
## UniqueAuthors3     0.3130    0.0368   8.50 < 2e-16 ***
## UniqueAuthors4     0.3952    0.0383  10.33 < 2e-16 ***
## UniqueAuthors5     0.4662    0.0354  13.18 < 2e-16 ***
## Year1997          -0.2076    0.2003  -1.04  0.30018
## Year1998          -0.3927    0.1795  -2.19  0.02877 *
## Year1999          -0.2565    0.1670  -1.54  0.12478
```

```

## Year2000          -0.2598      0.1764    -1.47    0.14107
## Year2001          -0.3147      0.1644    -1.91    0.05568 .
## Year2002          -0.5222      0.1510    -3.46    0.00055 ***
## Year2003          -0.5247      0.1440    -3.64    0.00027 ***
## Year2004          -0.4856      0.1470    -3.30    0.00097 ***
## Year2005          -0.6426      0.1405    -4.57    5.1e-06 ***
## Year2006          -0.6184      0.1407    -4.39    1.2e-05 ***
## Year2007          -0.4655      0.1369    -3.40    0.00068 ***
## Year2008          -0.4841      0.1367    -3.54    0.00041 ***
## Year2009          -0.4875      0.1349    -3.61    0.00031 ***
## Year2010          -0.4317      0.1357    -3.18    0.00149 **
## Year2011          -0.4538      0.1354    -3.35    0.00081 ***
## Year2012          -0.3915      0.1364    -2.87    0.00415 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.116, Adjusted R-squared:  0.108
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 187 weights are ~= 1. The remaining 2082 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0138 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          4.41e-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 1.211 1          1.101
## LastAuthorFemale 1.180 1          1.086
## Year          1.092 16          1.003

```

Residuals from first and last author



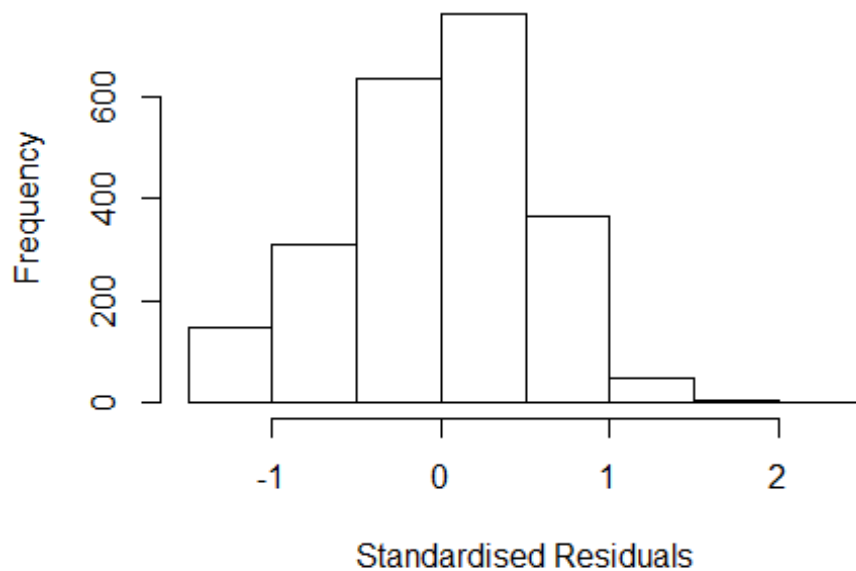
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4667 -0.3882 0.0254 0.3777 2.3862
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3940 0.1393 10.01 < 2e-16 ***
## FirstAuthorFemale1 0.0727 0.0275 2.64 0.00826 **
## LastAuthorFemale1 -0.0252 0.0265 -0.95 0.34235
## Year1997 -0.1600 0.2158 -0.74 0.45837
## Year1998 -0.3268 0.1962 -1.67 0.09592 .
## Year1999 -0.2113 0.1713 -1.23 0.21750
## Year2000 -0.2215 0.1800 -1.23 0.21871
## Year2001 -0.2164 0.1658 -1.31 0.19197
## Year2002 -0.4403 0.1556 -2.83 0.00470 **
## Year2003 -0.3986 0.1506 -2.65 0.00817 **
## Year2004 -0.4042 0.1540 -2.62 0.00873 **
## Year2005 -0.5273 0.1482 -3.56 0.00038 ***
```

```

## Year2006          -0.4845      0.1469   -3.30  0.00098 ***
## Year2007          -0.3460      0.1435   -2.41  0.01599 *
## Year2008          -0.3659      0.1434   -2.55  0.01078 *
## Year2009          -0.3983      0.1416   -2.81  0.00497 **
## Year2010          -0.3316      0.1431   -2.32  0.02063 *
## Year2011          -0.3093      0.1421   -2.18  0.02956 *
## Year2012          -0.2240      0.1425   -1.57  0.11626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.028, Adjusted R-squared:  0.0202
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2069 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0318 0.8650 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      4.41e-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 1.056 1      1.028
## Year      1.056 16      1.002

```

Residuals from first author



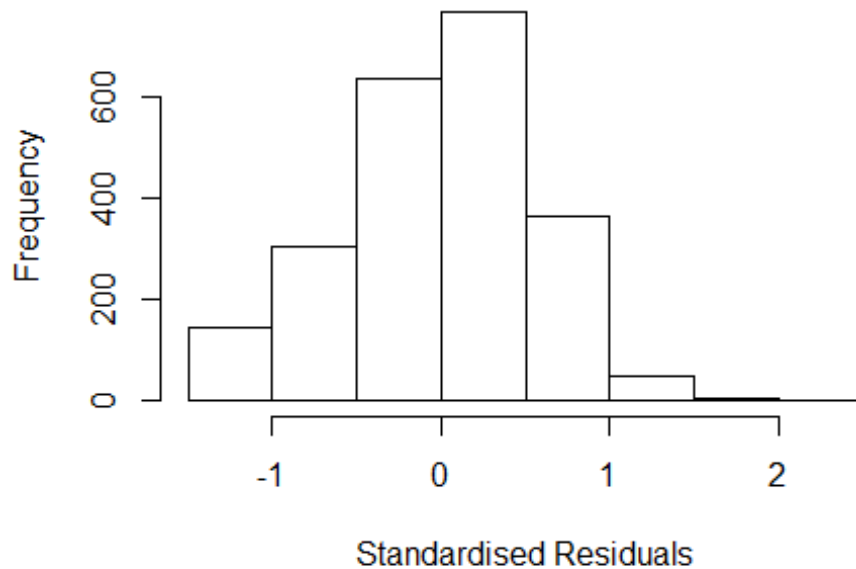
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4481 -0.3891  0.0228  0.3803  2.4030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3850     0.1385   10.00  <2e-16 ***
## FirstAuthorFemale1  0.0631     0.0257    2.46  0.0140 *
## Year1997         -0.1601     0.2148   -0.75  0.4560
## Year1998         -0.3215     0.1959   -1.64  0.1010
## Year1999         -0.2056     0.1709   -1.20  0.2290
## Year2000         -0.2170     0.1794   -1.21  0.2265
## Year2001         -0.2141     0.1657   -1.29  0.1965
## Year2002         -0.4378     0.1555   -2.81  0.0049 **
## Year2003         -0.3968     0.1504   -2.64  0.0084 **
## Year2004         -0.4011     0.1539   -2.61  0.0092 **
## Year2005         -0.5245     0.1479   -3.55  0.0004 ***
## Year2006         -0.4836     0.1468   -3.30  0.0010 ***
```

```

## Year2007          -0.3441      0.1433   -2.40    0.0165 *
## Year2008          -0.3629      0.1431   -2.54    0.0113 *
## Year2009          -0.3962      0.1415   -2.80    0.0052 **
## Year2010          -0.3297      0.1430   -2.31    0.0212 *
## Year2011          -0.3067      0.1419   -2.16    0.0307 *
## Year2012          -0.2203      0.1423   -1.55    0.1216
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.0202
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 207 weights are ~= 1. The remaining 2062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0275 0.8650 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      4.41e-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 1.031 1      1.015
## Year      1.031 16      1.001

```


Residuals from last author



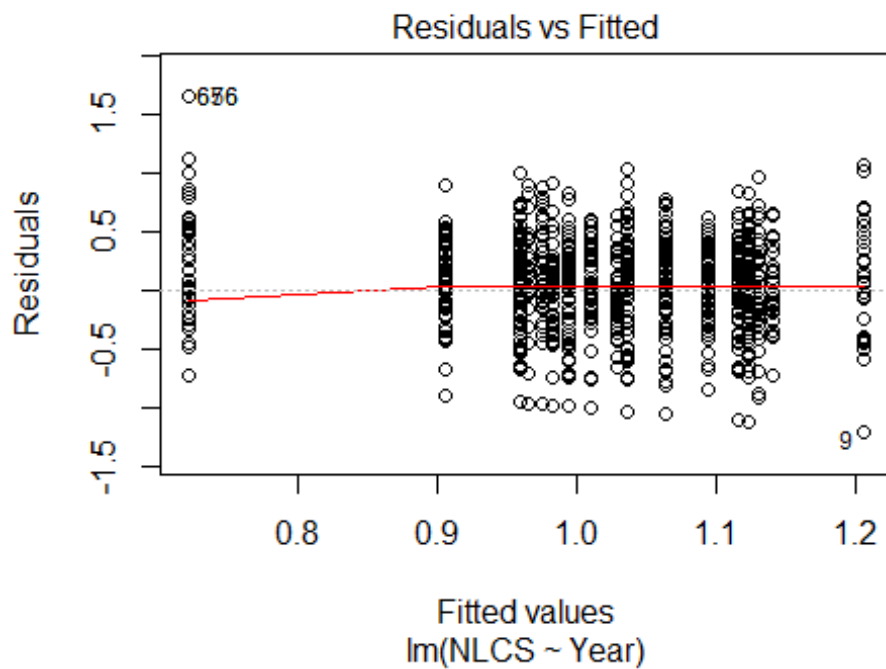
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4169 -0.3866 0.0261 0.3839 2.4283
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.41694 0.14013 10.11 < 2e-16 ***
## LastAuthorFemale1 0.00143 0.02479 0.06 0.95395
## Year1997 -0.17132 0.21648 -0.79 0.42881
## Year1998 -0.31225 0.19532 -1.60 0.11004
## Year1999 -0.22024 0.17325 -1.27 0.20377
## Year2000 -0.21670 0.18004 -1.20 0.22887
## Year2001 -0.22129 0.16810 -1.32 0.18817
## Year2002 -0.43645 0.15619 -2.79 0.00524 **
## Year2003 -0.39495 0.15144 -2.61 0.00917 **
## Year2004 -0.39930 0.15539 -2.57 0.01025 *
## Year2005 -0.51698 0.14889 -3.47 0.00053 ***
## Year2006 -0.47896 0.14771 -3.24 0.00120 **
```

```

## Year2007          -0.33824      0.14430      -2.34      0.01916 *
## Year2008          -0.35889      0.14415      -2.49      0.01286 *
## Year2009          -0.39405      0.14260      -2.76      0.00577 **
## Year2010          -0.32190      0.14390      -2.24      0.02538 *
## Year2011          -0.30075      0.14285      -2.11      0.03538 *
## Year2012          -0.21284      0.14327      -1.49      0.13753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0246, Adjusted R-squared:  0.0172
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 197 weights are ~= 1. The remaining 2072 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0223 0.8670 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      4.41e-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: 2269"
## [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
## 147 153 99 73 154 122 122 63 80 73 91 106 105 110 119
## 2011 2012
## 113 142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 35 45 25 30 57 39 77 37 55 54 60 71 77 78 75
## 2011 2012

```

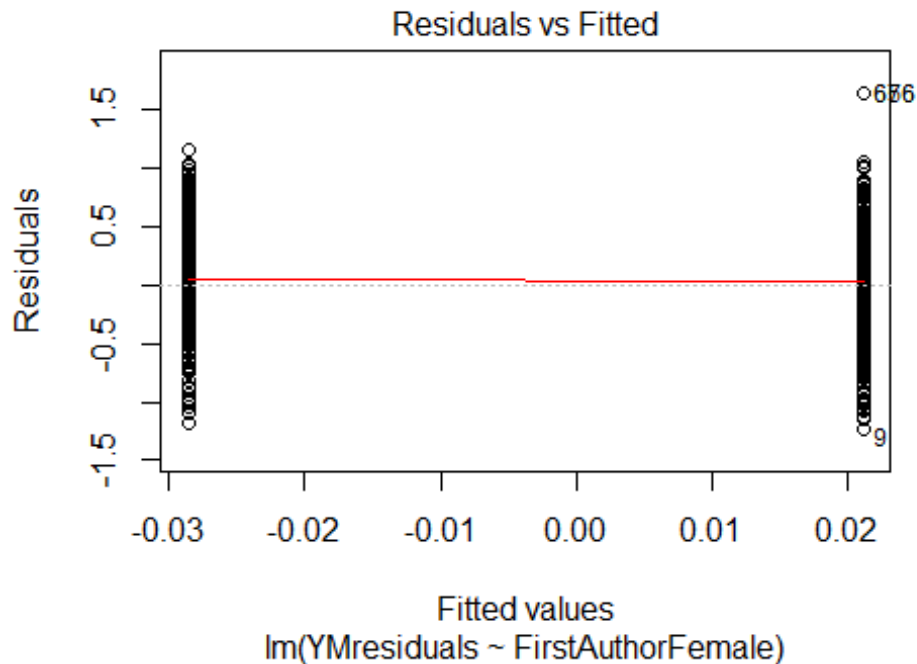
```
## 79 97
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 32 37 17 26 47 32 61 28 45 43 49 60 69 69 70
## 2011 2012
## 69 90
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 60, df = 16, p-value = 5e-07
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.3, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 5.22391417111534"
## [1] "Male first author team size 2018 geometric mean: 4.69963693871122"

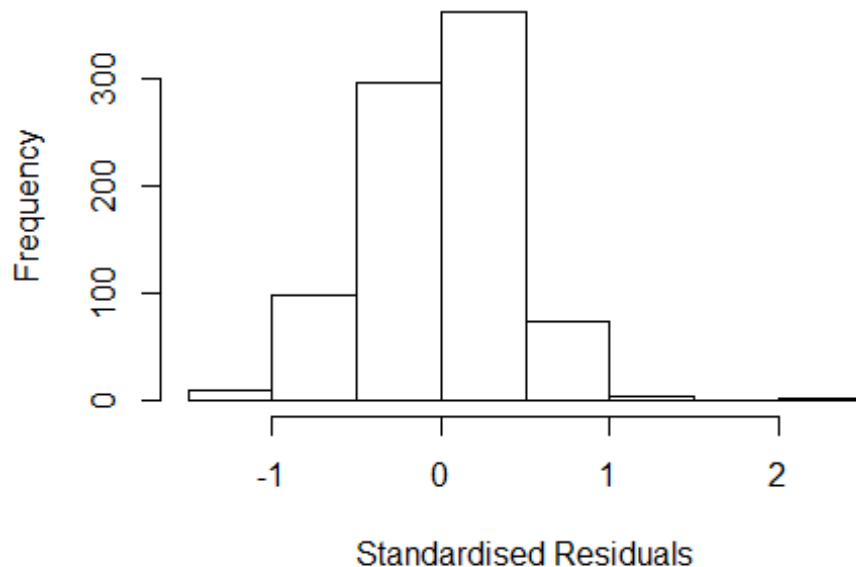
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 700, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.94995840288192"
## [1] "Male last author team size 2018 geometric mean: 5.06111842805624"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 600, 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	1.036
LastAuthorFemale	1.101	1	1.049
UniqueAuthors	1.600	4	1.061
Year	1.743	16	1.018

Residuals from first and last author and team size



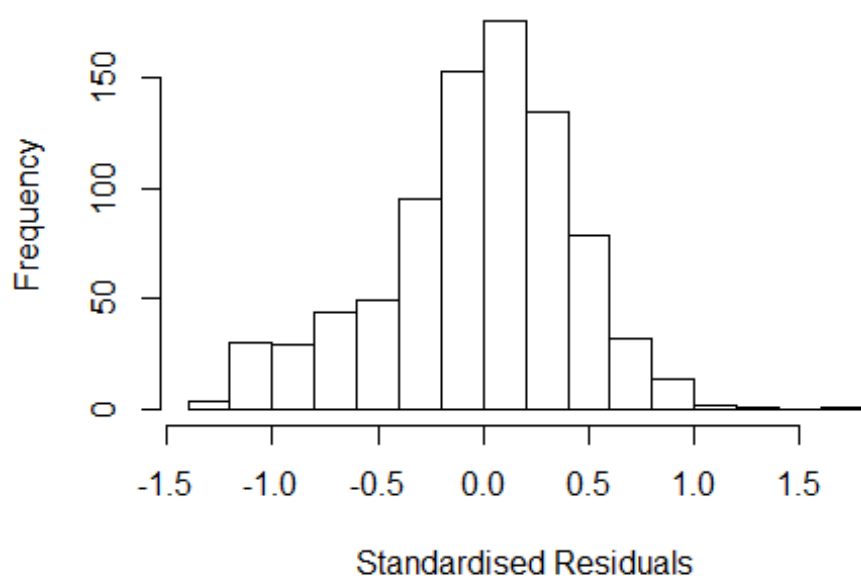
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2978 -0.2680 0.0187 0.2604 2.0438
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8751 0.1087 8.05 2.9e-15 ***
## FirstAuthorFemale1 -0.0487 0.0296 -1.64 0.10051
## LastAuthorFemale1 -0.0621 0.0346 -1.79 0.07353 .
## UniqueAuthors2 0.2866 0.0968 2.96 0.00314 **
## UniqueAuthors3 0.4227 0.0873 4.84 1.5e-06 ***
## UniqueAuthors4 0.3767 0.0915 4.12 4.3e-05 ***
## UniqueAuthors5 0.6061 0.0828 7.32 6.0e-13 ***
## Year1997 -0.2900 0.1376 -2.11 0.03534 *
## Year1998 -0.2207 0.1311 -1.68 0.09270 .
## Year1999 -0.1414 0.1009 -1.40 0.16148
```

```

## Year2000          -0.5409      0.1488    -3.63  0.00030 ***
## Year2001          -0.1109      0.1101    -1.01  0.31443
## Year2002          -0.1616      0.0984    -1.64  0.10095
## Year2003          -0.2401      0.1006    -2.39  0.01717 *
## Year2004          -0.3962      0.1017    -3.90  0.00011 ***
## Year2005          -0.2904      0.0941    -3.08  0.00211 **
## Year2006          -0.2609      0.1064    -2.45  0.01442 *
## Year2007          -0.2751      0.0947    -2.91  0.00377 **
## Year2008          -0.2265      0.0959    -2.36  0.01843 *
## Year2009          -0.2682      0.0972    -2.76  0.00593 **
## Year2010          -0.1934      0.0930    -2.08  0.03801 *
## Year2011          -0.2013      0.0994    -2.03  0.04314 *
## Year2012          -0.3202      0.0933    -3.43  0.00063 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.237, Adjusted R-squared:  0.216
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## observation 146 is an outlier with |weight| = 0 ( < 0.00012);
## 79 weights are ~= 1. The remaining 764 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.211  0.848  0.942  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          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 1.081 1          1.040
## LastAuthorFemale  1.086 1          1.042
## Year              1.170 16          1.005

```

Residuals from first and last author



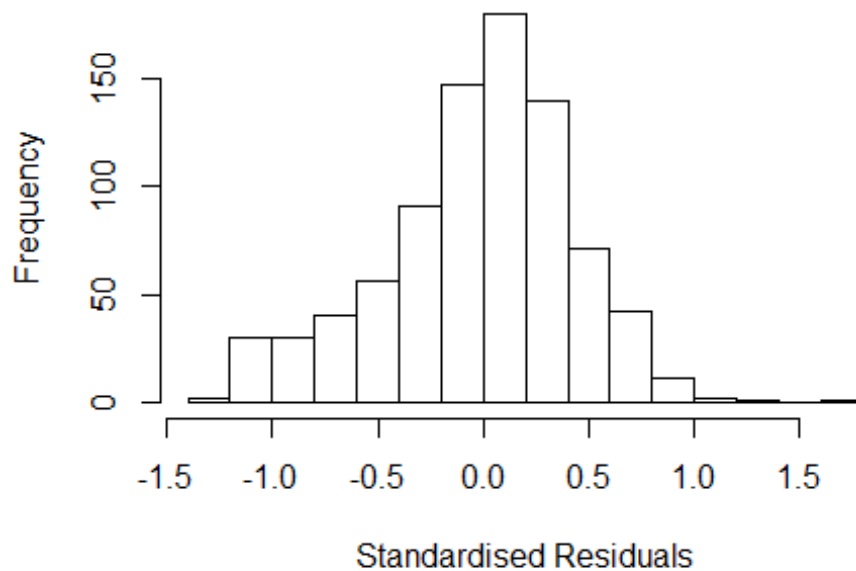
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2725 -0.2800  0.0262  0.2744  1.7245
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2725     0.0889   14.31  <2e-16 ***
## FirstAuthorFemale1 -0.0322     0.0320   -1.01   0.3134
## LastAuthorFemale1  -0.0998     0.0380   -2.63   0.0088 **
## Year1997          -0.2371     0.1406   -1.69   0.0921 .
## Year1998          -0.1571     0.1420   -1.11   0.2690
## Year1999          -0.0984     0.1067   -0.92   0.3571
## Year2000          -0.6191     0.1535   -4.03   6e-05 ***
## Year2001          -0.0264     0.1167   -0.23   0.8214
## Year2002          -0.0706     0.1034   -0.68   0.4946
## Year2003          -0.1723     0.1105   -1.56   0.1192
## Year2004          -0.3621     0.1127   -3.21   0.0014 **
## Year2005          -0.2060     0.0993   -2.07   0.0383 *
```

```

## Year2006          -0.1772      0.1095   -1.62    0.1059
## Year2007          -0.1413      0.0980   -1.44    0.1500
## Year2008          -0.1409      0.0992   -1.42    0.1558
## Year2009          -0.1712      0.1026   -1.67    0.0954 .
## Year2010          -0.0833      0.0991   -0.84    0.4005
## Year2011          -0.1224      0.1047   -1.17    0.2428
## Year2012          -0.2383      0.1046   -2.28    0.0230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0923, Adjusted R-squared:  0.0724
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 766 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0173 0.8420 0.9450 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      1.18e-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 1.075 1      1.037
## Year              1.075 16      1.002

```


Residuals from first author



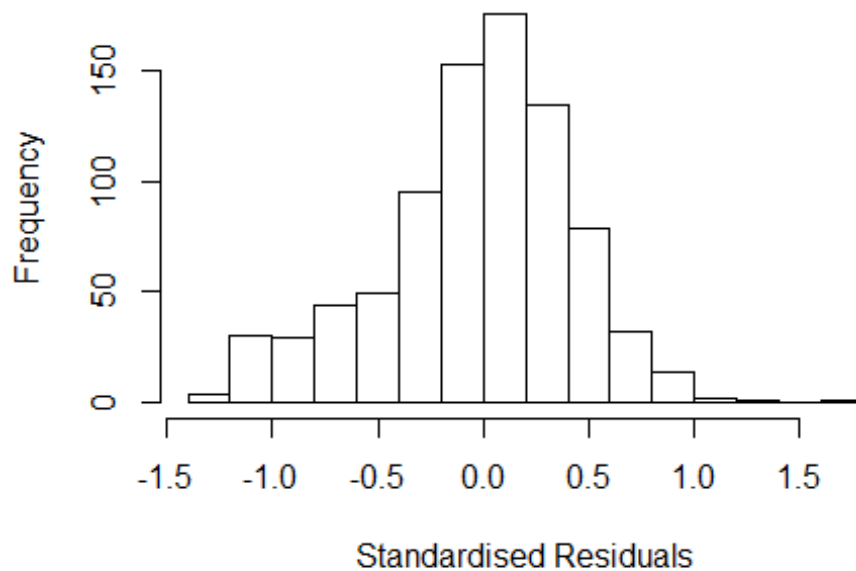
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2455 -0.2859  0.0283  0.2782  1.7474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2455     0.0897   13.88 < 2e-16 ***
## FirstAuthorFemale1 -0.0357     0.0321   -1.11  0.266
## Year1997          -0.2461     0.1418   -1.74  0.083 .
## Year1998          -0.1305     0.1426   -0.92  0.360
## Year1999          -0.0773     0.1074   -0.72  0.472
## Year2000          -0.6149     0.1543   -3.98 7.4e-05 ***
## Year2001          -0.0147     0.1181   -0.12  0.901
## Year2002          -0.0643     0.1048   -0.61  0.540
## Year2003          -0.1606     0.1120   -1.43  0.152
## Year2004          -0.3565     0.1152   -3.10  0.002 **
## Year2005          -0.2061     0.1009   -2.04  0.041 *
## Year2006          -0.1683     0.1092   -1.54  0.124
```

```

## Year2007          -0.1481      0.1000   -1.48    0.139
## Year2008          -0.1311      0.1009   -1.30    0.194
## Year2009          -0.1725      0.1038   -1.66    0.097 .
## Year2010          -0.0876      0.1007   -0.87    0.384
## Year2011          -0.1159      0.1065   -1.09    0.277
## Year2012          -0.2399      0.1068   -2.25    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.0832, Adjusted R-squared:  0.0643
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 781 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0163 0.8400 0.9480 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      1.18e-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 1.087 1          1.043
## Year            1.087 16          1.003

```

Residuals from last author



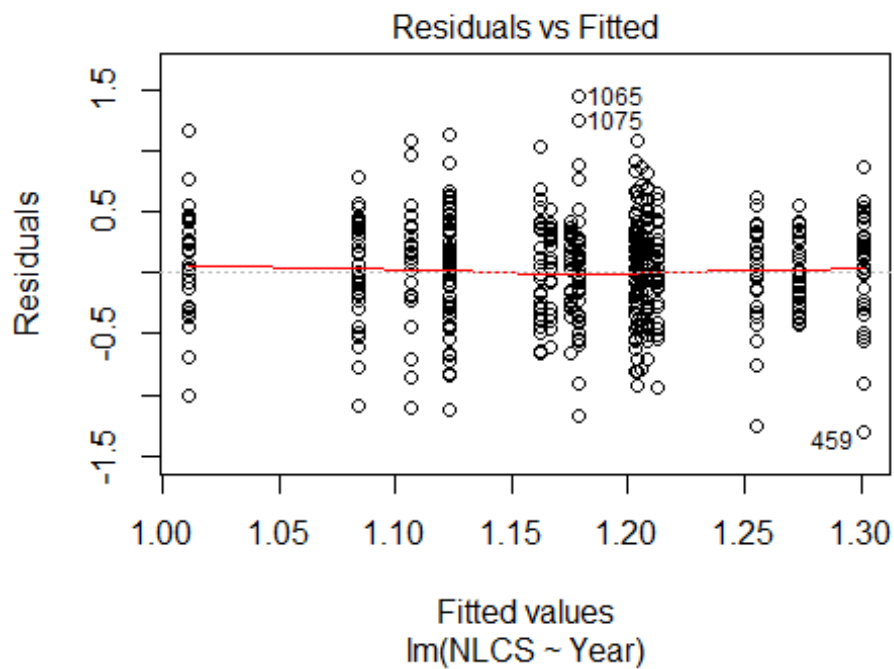
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2639 -0.2738  0.0156  0.2735  1.7331
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2639     0.0881  14.35  < 2e-16 ***
## LastAuthorFemale1 -0.1013     0.0381  -2.66  0.00801 **
## Year1997         -0.2362     0.1417  -1.67  0.09589 .
## Year1998         -0.1589     0.1422  -1.12  0.26422
## Year1999         -0.1005     0.1067  -0.94  0.34676
## Year2000         -0.6190     0.1531  -4.04  5.7e-05 ***
## Year2001         -0.0335     0.1176  -0.29  0.77568
## Year2002         -0.0746     0.1028  -0.73  0.46841
## Year2003         -0.1805     0.1094  -1.65  0.09913 .
## Year2004         -0.3691     0.1113  -3.32  0.00096 ***
## Year2005         -0.2098     0.0989  -2.12  0.03413 *
## Year2006         -0.1821     0.1091  -1.67  0.09563 .
```

```

## Year2007          -0.1459      0.0976   -1.50  0.13506
## Year2008          -0.1495      0.0983   -1.52  0.12851
## Year2009          -0.1743      0.1024   -1.70  0.08905 .
## Year2010          -0.0904      0.0986   -0.92  0.35981
## Year2011          -0.1275      0.1045   -1.22  0.22278
## Year2012          -0.2442      0.1043   -2.34  0.01945 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0904, Adjusted R-squared:  0.0717
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 766 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.015  0.836  0.948  0.882  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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: 844"
## [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
##   70   46   66   55   74   58   53   44   52   48   41   54   68   74   47
## 2011 2012
##   54   54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   24   32   25   46   32   33   28   29   25   27   38   41   45   37
## 2011 2012

```

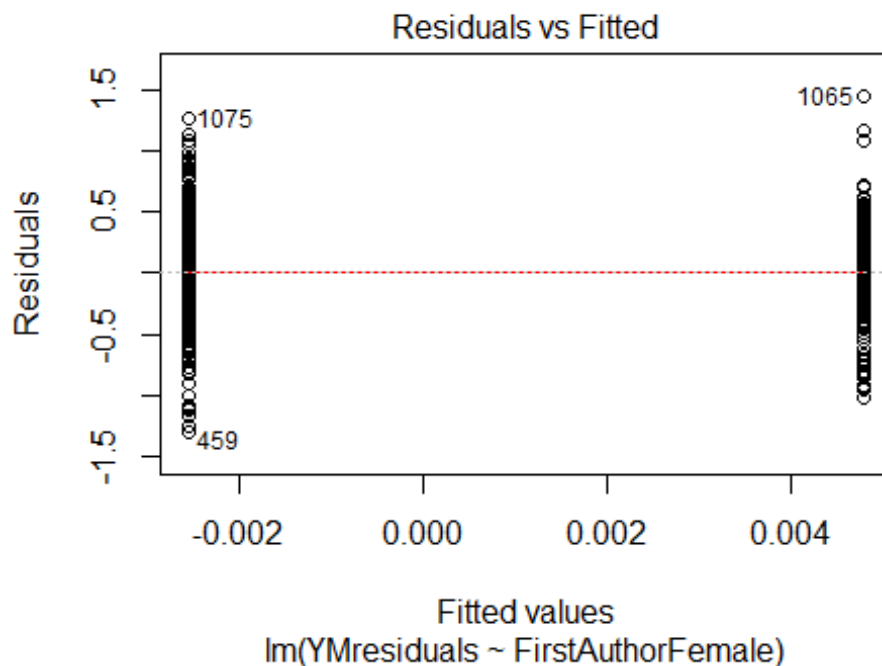
```
## 39 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 31 23 28 24 43 26 32 26 29 22 22 31 39 43 35
## 2011 2012
## 34 34
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, 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: 3.76854777871428"
## [1] "Male first author team size 2018 geometric mean: 3.75097329182709"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

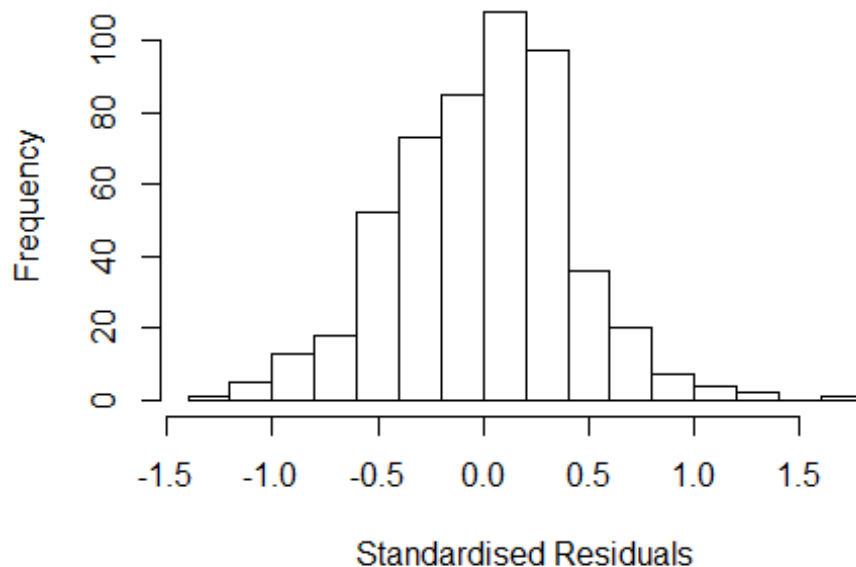
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 150, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77454961710008"
## [1] "Male last author team size 2018 geometric mean: 3.7503715663724"

## 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 = 0.7
## 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 1.235  1      1.111
## LastAuthorFemale  1.175  1      1.084
## UniqueAuthors    1.855  4      1.080
## Year              2.204 16      1.025
```

Residuals from first and last author and team size



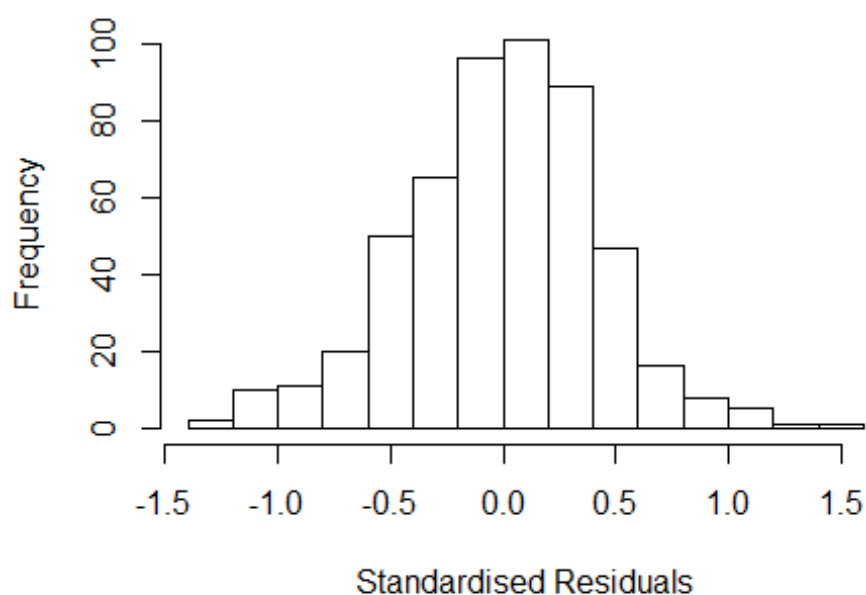
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2387 -0.2871 0.0192 0.2525 1.6662
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86993 0.13948 6.24 9.5e-10 ***
## FirstAuthorFemale1 -0.03031 0.03799 -0.80 0.42544
## LastAuthorFemale1 -0.03497 0.04557 -0.77 0.44312
## UniqueAuthors2 0.28951 0.14457 2.00 0.04577 *
## UniqueAuthors3 0.25277 0.13794 1.83 0.06748 .
## UniqueAuthors4 0.33475 0.13550 2.47 0.01383 *
## UniqueAuthors5 0.46667 0.12923 3.61 0.00034 ***
## Year1997 -0.05645 0.10437 -0.54 0.58884
## Year1998 -0.01616 0.11598 -0.14 0.88923
## Year1999 0.00648 0.11633 0.06 0.95561
```

```

## Year2000      -0.00298    0.10339   -0.03  0.97698
## Year2001      0.04856    0.10100    0.48  0.63088
## Year2002      0.11603    0.10774    1.08  0.28202
## Year2003      0.04384    0.09426    0.47  0.64208
## Year2004     -0.03722    0.11598   -0.32  0.74837
## Year2005     -0.06152    0.09689   -0.63  0.52575
## Year2006     -0.10921    0.11432   -0.96  0.33986
## Year2007     -0.03409    0.10095   -0.34  0.73573
## Year2008     -0.02673    0.10422   -0.26  0.79770
## Year2009     -0.13428    0.10827   -1.24  0.21549
## Year2010     -0.13033    0.10606   -1.23  0.21970
## Year2011     -0.13980    0.11811   -1.18  0.23710
## Year2012     -0.15802    0.11950   -1.32  0.18666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0659
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 479 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0474 0.8840 0.9510 0.9030 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          1.92e-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 1.160 1          1.077
## LastAuthorFemale 1.134 1          1.065
## Year      1.314 16          1.009

```


Residuals from first and last author



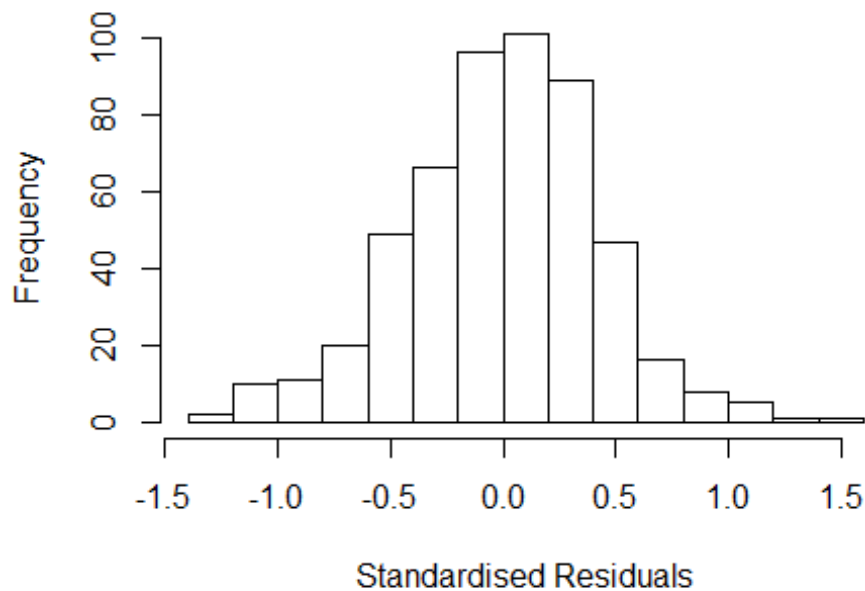
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35407 -0.29125  0.00944  0.28734  1.48471
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.19205    0.08562   13.92  <2e-16 ***
## FirstAuthorFemale1  0.00687    0.03842    0.18    0.86
## LastAuthorFemale1 -0.00356    0.04941   -0.07    0.94
## Year1997          -0.03148    0.11141   -0.28    0.78
## Year1998          -0.03598    0.11774   -0.31    0.76
## Year1999           0.02337    0.11714    0.20    0.84
## Year2000           0.04398    0.10408    0.42    0.67
## Year2001           0.07393    0.10680    0.69    0.49
## Year2002           0.16202    0.11155    1.45    0.15
## Year2003           0.07265    0.10521    0.69    0.49
## Year2004          -0.04540    0.13735   -0.33    0.74
## Year2005          -0.00598    0.10513   -0.06    0.95
```

```

## Year2006      -0.05826    0.12731   -0.46    0.65
## Year2007      0.02307    0.10658    0.22    0.83
## Year2008      0.01813    0.10857    0.17    0.87
## Year2009     -0.06290    0.11411   -0.55    0.58
## Year2010     -0.05006    0.11645   -0.43    0.67
## Year2011     -0.10339    0.11720   -0.88    0.38
## Year2012     -0.14485    0.12584   -1.15    0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0292, Adjusted R-squared:  -0.00555
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.884  0.947  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      1.92e-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 1.156 1      1.075
## Year              1.156 16      1.005

```

Residuals from first author



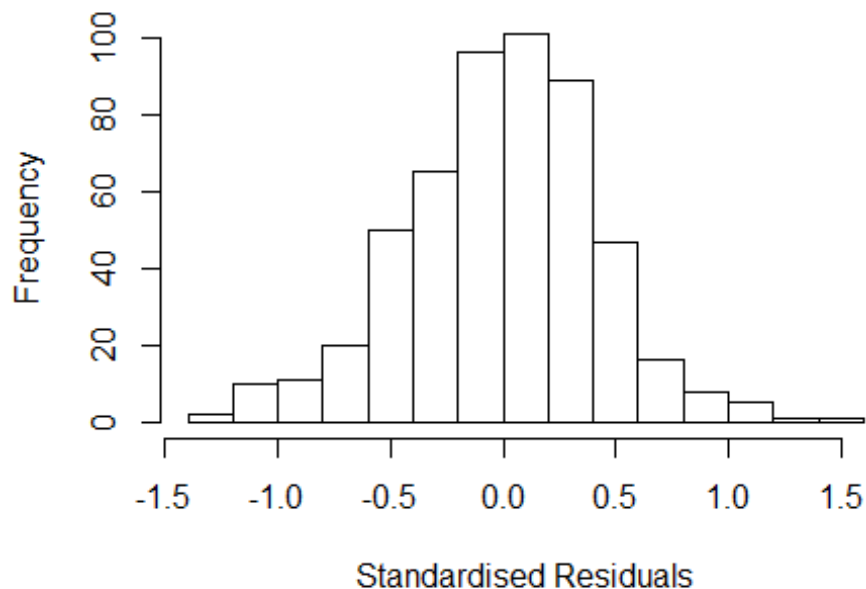
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35387 -0.29054 0.00814 0.28799 1.48180
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19154 0.08594 13.86 <2e-16 ***
## FirstAuthorFemale1 0.00665 0.03842 0.17 0.86
## Year1997 -0.03159 0.11143 -0.28 0.78
## Year1998 -0.03579 0.11793 -0.30 0.76
## Year1999 0.02352 0.11713 0.20 0.84
## Year2000 0.04391 0.10411 0.42 0.67
## Year2001 0.07393 0.10664 0.69 0.49
## Year2002 0.16233 0.11168 1.45 0.15
## Year2003 0.07208 0.10435 0.69 0.49
## Year2004 -0.04501 0.13754 -0.33 0.74
## Year2005 -0.00566 0.10538 -0.05 0.96
## Year2006 -0.05825 0.12746 -0.46 0.65
```

```

## Year2007          0.02305    0.10664    0.22    0.83
## Year2008          0.01764    0.10728    0.16    0.87
## Year2009         -0.06288    0.11409   -0.55    0.58
## Year2010         -0.05000    0.11657   -0.43    0.67
## Year2011         -0.10317    0.11745   -0.88    0.38
## Year2012         -0.14458    0.12588   -1.15    0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0292, Adjusted R-squared:  -0.00354
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 471 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.169  0.882  0.947  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      1.92e-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 1.128 1          1.062
## Year            1.128 16          1.004

```

Residuals from last author



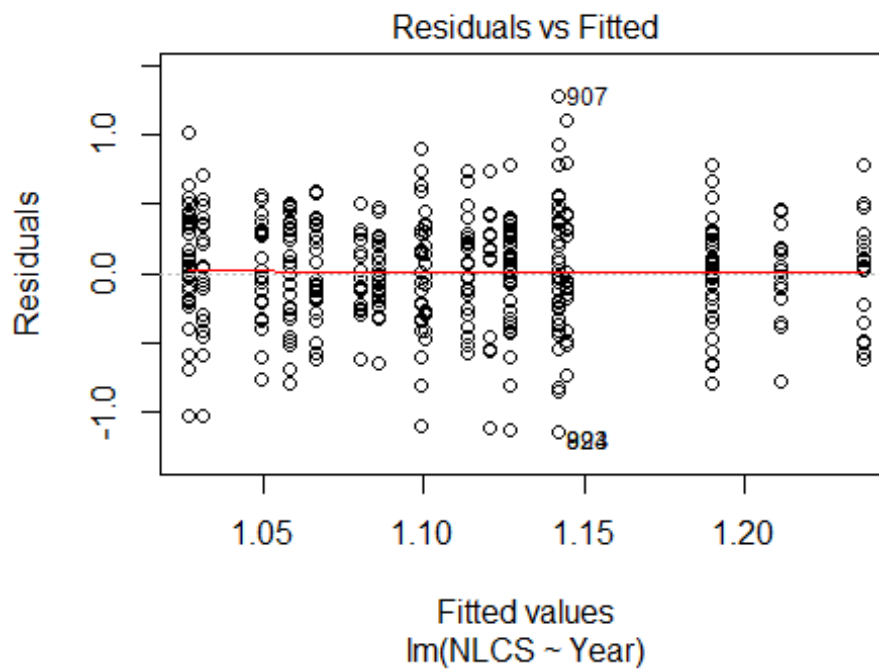
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.356 -0.287 0.011 0.287 1.488
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19403 0.08397 14.22 <2e-16 ***
## LastAuthorFemale1 -0.00312 0.04934 -0.06 0.95
## Year1997 -0.03058 0.11137 -0.27 0.78
## Year1998 -0.03558 0.11770 -0.30 0.76
## Year1999 0.02394 0.11734 0.20 0.84
## Year2000 0.04471 0.10408 0.43 0.67
## Year2001 0.07459 0.10681 0.70 0.49
## Year2002 0.16228 0.11154 1.45 0.15
## Year2003 0.07337 0.10547 0.70 0.49
## Year2004 -0.04395 0.13689 -0.32 0.75
## Year2005 -0.00470 0.10508 -0.04 0.96
## Year2006 -0.05854 0.12709 -0.46 0.65
```

```

## Year2007          0.02242      0.10612      0.21      0.83
## Year2008          0.01932      0.10933      0.18      0.86
## Year2009         -0.06231      0.11398     -0.55      0.58
## Year2010         -0.04885      0.11570     -0.42      0.67
## Year2011         -0.10429      0.11679     -0.89      0.37
## Year2012         -0.14397      0.12581     -1.14      0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0291, Adjusted R-squared:  -0.00366
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 471 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.167  0.882  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      1.92e-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: 522"
## [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
##   44   50   36   43   39   38   53   35   36   43   37   49   42   47   59
## 2011 2012
##   61   54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   22   17   28   19   14   22   21   17   26   25   24   23   33   38
## 2011 2012

```

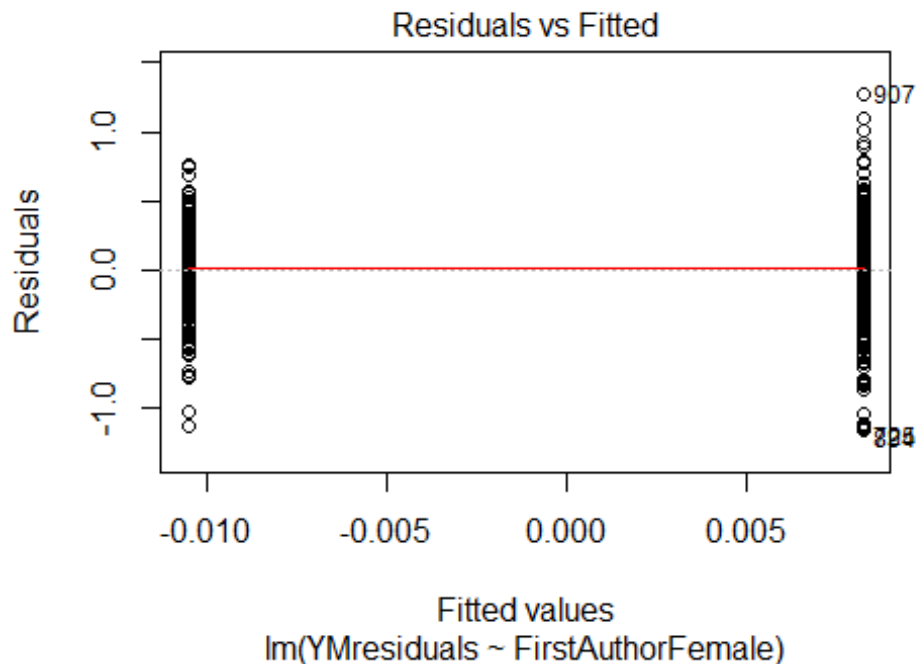
```
## 35 37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 19 16 27 17 12 20 20 13 20 20 20 20 24 33
## 2011 2012
## 30 32
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```



```
##
## 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: 5.62311984961588"
## [1] "Male first author team size 2018 geometric mean: 3.44982049539645"
## 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 = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.25977486795328"
## [1] "Male last author team size 2018 geometric mean: 4.44255660148963"

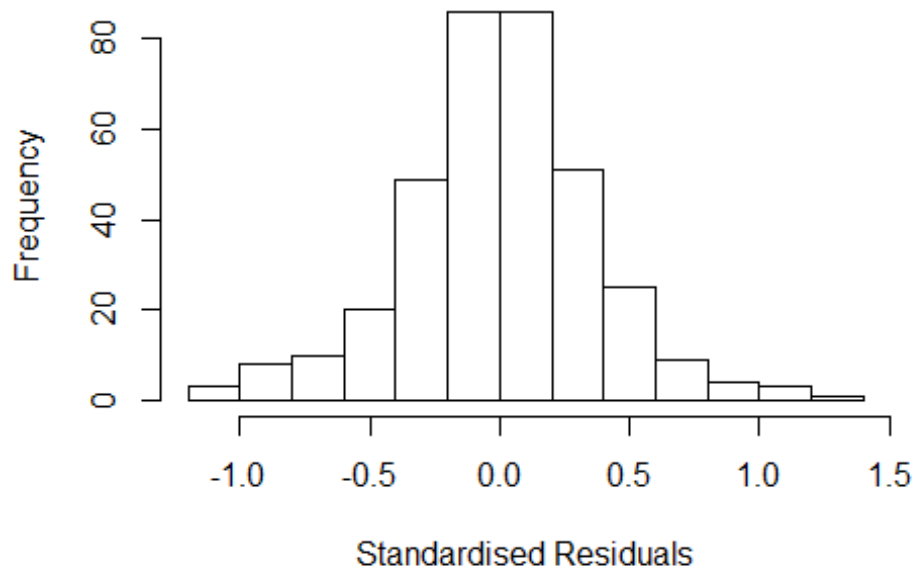
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.5
## 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	1.538	1	1.240
LastAuthorFemale	1.393	1	1.180
UniqueAuthors	2.274	4	1.108
Year	3.255	16	1.038

Residuals from first and last author and team size



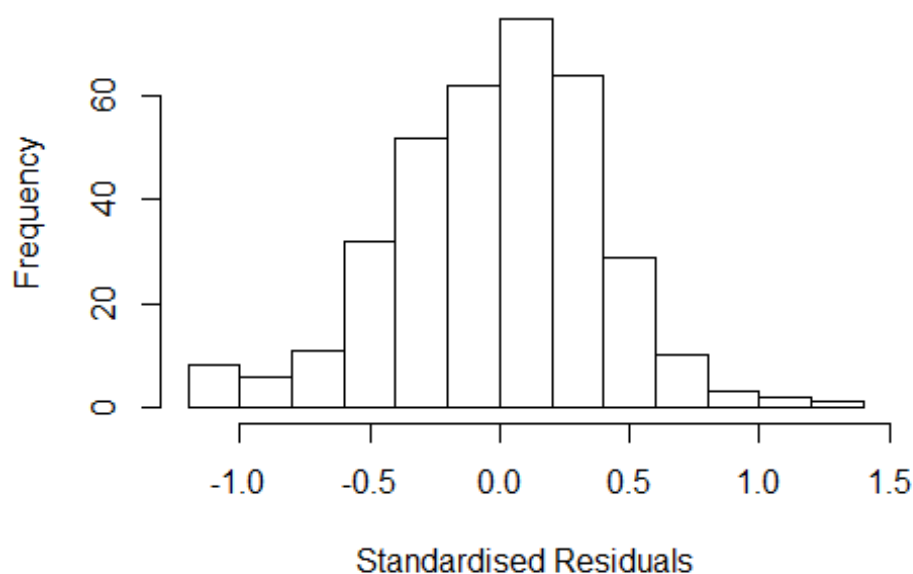
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11782 -0.20582 0.00298 0.20508 1.29648
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04059 0.17874 5.82 1.4e-08 ***
## FirstAuthorFemale1 -0.01438 0.04194 -0.34 0.73
## LastAuthorFemale1 -0.00868 0.04283 -0.20 0.84
## UniqueAuthors2 -0.01140 0.10311 -0.11 0.91
## UniqueAuthors3 0.04776 0.09431 0.51 0.61
## UniqueAuthors4 0.16425 0.09671 1.70 0.09 .
## UniqueAuthors5 0.36857 0.09087 4.06 6.2e-05 ***
## Year1997 -0.03630 0.19800 -0.18 0.85
## Year1998 0.12342 0.16338 0.76 0.45
## Year1999 -0.09319 0.16415 -0.57 0.57
```

```

## Year2000      -0.09518    0.19819   -0.48    0.63
## Year2001      -0.00846    0.17060   -0.05    0.96
## Year2002      -0.05445    0.15910   -0.34    0.73
## Year2003      -0.08213    0.15991   -0.51    0.61
## Year2004       0.16799    0.19262    0.87    0.38
## Year2005      -0.07609    0.16295   -0.47    0.64
## Year2006      -0.13295    0.15552   -0.85    0.39
## Year2007      -0.15878    0.16627   -0.95    0.34
## Year2008      -0.11231    0.17146   -0.66    0.51
## Year2009      -0.01242    0.16053   -0.08    0.94
## Year2010      -0.06452    0.15818   -0.41    0.68
## Year2011      -0.20386    0.16357   -1.25    0.21
## Year2012      -0.07263    0.17754   -0.41    0.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.173, Adjusted R-squared:  0.118
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 323 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  0.876  0.959  0.897  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      2.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.280 1      1.131
## LastAuthorFemale  1.438 1      1.199
## Year              1.810 16      1.019

```

Residuals from first and last author



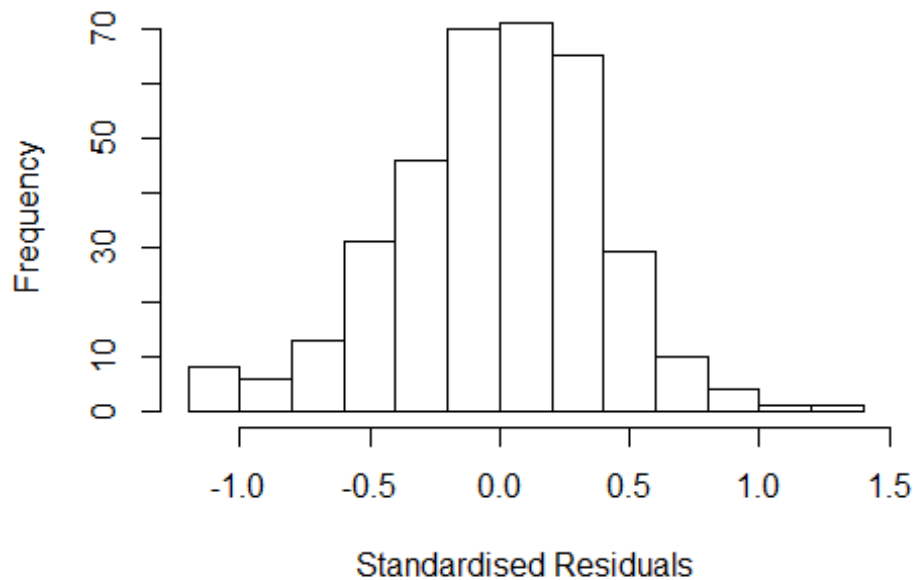
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1967 -0.2576  0.0101  0.2553  1.2879
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1967     0.1749   6.84 3.7e-11 ***
## FirstAuthorFemale1 -0.0317     0.0443  -0.71   0.48
## LastAuthorFemale1 -0.0196     0.0480  -0.41   0.68
## Year1997          -0.1018     0.2029  -0.50   0.62
## Year1998           0.0569     0.1815   0.31   0.75
## Year1999          -0.1162     0.1855  -0.63   0.53
## Year2000          -0.1751     0.2038  -0.86   0.39
## Year2001          -0.0772     0.1906  -0.40   0.69
## Year2002          -0.0787     0.1775  -0.44   0.66
## Year2003          -0.1020     0.1788  -0.57   0.57
## Year2004           0.1185     0.2062   0.57   0.57
## Year2005          -0.0208     0.1932  -0.11   0.91
```

```

## Year2006          -0.1021      0.1796   -0.57      0.57
## Year2007          -0.0989      0.1875   -0.53      0.60
## Year2008          -0.0961      0.1883   -0.51      0.61
## Year2009           0.0639      0.1774    0.36      0.72
## Year2010          -0.0272      0.1770   -0.15      0.88
## Year2011          -0.1392      0.1850   -0.75      0.45
## Year2012          -0.0450      0.1973   -0.23      0.82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.0341, Adjusted R-squared:  -0.0177
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 322 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.237  0.876  0.948  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      2.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.282 1      1.132
## Year              1.282 16      1.008

```

Residuals from first author



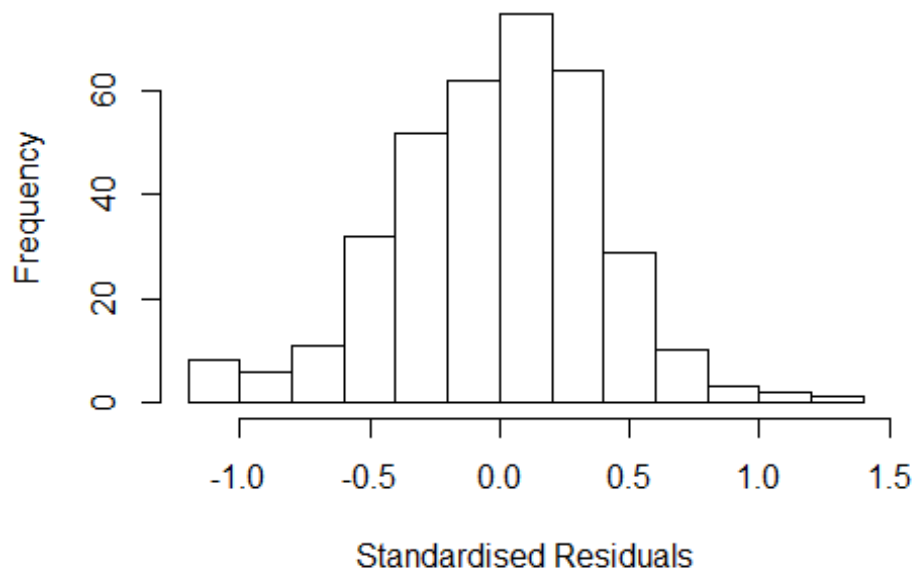
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1878 -0.2513 0.0139 0.2587 1.2755
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1878 0.1728 6.87 3.1e-11 ***
## FirstAuthorFemale1 -0.0322 0.0444 -0.72 0.47
## Year1997 -0.0975 0.2032 -0.48 0.63
## Year1998 0.0541 0.1826 0.30 0.77
## Year1999 -0.1144 0.1863 -0.61 0.54
## Year2000 -0.1714 0.2032 -0.84 0.40
## Year2001 -0.0720 0.1909 -0.38 0.71
## Year2002 -0.0772 0.1783 -0.43 0.67
## Year2003 -0.0992 0.1795 -0.55 0.58
## Year2004 0.1219 0.2060 0.59 0.55
## Year2005 -0.0161 0.1917 -0.08 0.93
## Year2006 -0.0965 0.1791 -0.54 0.59
```

```

## Year2007          -0.0975      0.1882    -0.52      0.60
## Year2008          -0.0974      0.1889    -0.52      0.61
## Year2009           0.0653      0.1784     0.37      0.71
## Year2010          -0.0221      0.1771    -0.12      0.90
## Year2011          -0.1361      0.1851    -0.74      0.46
## Year2012          -0.0433      0.1976    -0.22      0.83
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.033, Adjusted R-squared:  -0.0158
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 318 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.877   0.946   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      2.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.425 1      1.194
## Year              1.425 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
## -1.1703 -0.2590 0.0118 0.2624 1.3041
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17034 0.16370 7.15 5.5e-12 ***
## LastAuthorFemale1 -0.02035 0.04785 -0.43 0.67
## Year1997 -0.09234 0.20060 -0.46 0.65
## Year1998 0.06740 0.17790 0.38 0.71
## Year1999 -0.10242 0.18010 -0.57 0.57
## Year2000 -0.16232 0.19761 -0.82 0.41
## Year2001 -0.06477 0.18584 -0.35 0.73
## Year2002 -0.06704 0.17280 -0.39 0.70
## Year2003 -0.08932 0.17317 -0.52 0.61
## Year2004 0.12638 0.20310 0.62 0.53
## Year2005 -0.00413 0.18436 -0.02 0.98
## Year2006 -0.09021 0.17565 -0.51 0.61
```

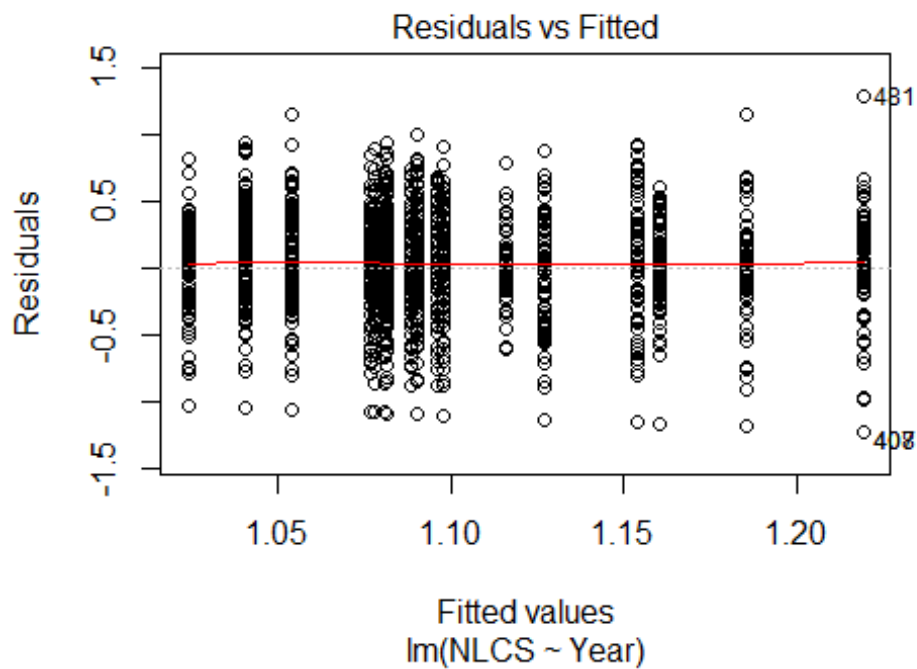
```

## Year2007          -0.08687      0.18346    -0.47      0.64
## Year2008          -0.08970      0.18615    -0.48      0.63
## Year2009           0.07314      0.17295     0.42      0.67
## Year2010          -0.01647      0.17291    -0.10      0.92
## Year2011          -0.13047      0.18120    -0.72      0.47
## Year2012          -0.03410      0.19418    -0.18      0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.032, Adjusted R-squared:  -0.0168
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 318 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.242  0.879   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      2.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 355"
## [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
## 156 160 157 103 200 139 139 102 147 116 145 148 147 147 159
## 2011 2012
## 190 210
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 59 76 74 55 71 60 88 75 92 80 101 92 94 91 107
## 2011 2012

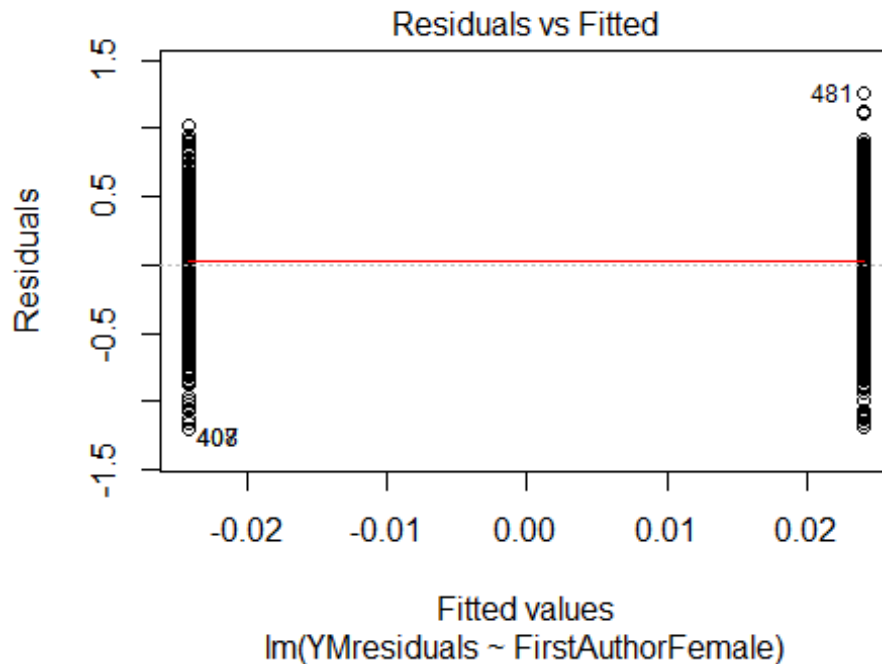
```



```
## 125 144
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 66 63 52 67 53 77 56 81 69 93 81 78 83 92
## 2011 2012
## 108 132
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 47, df = 16, p-value = 7e-05
```

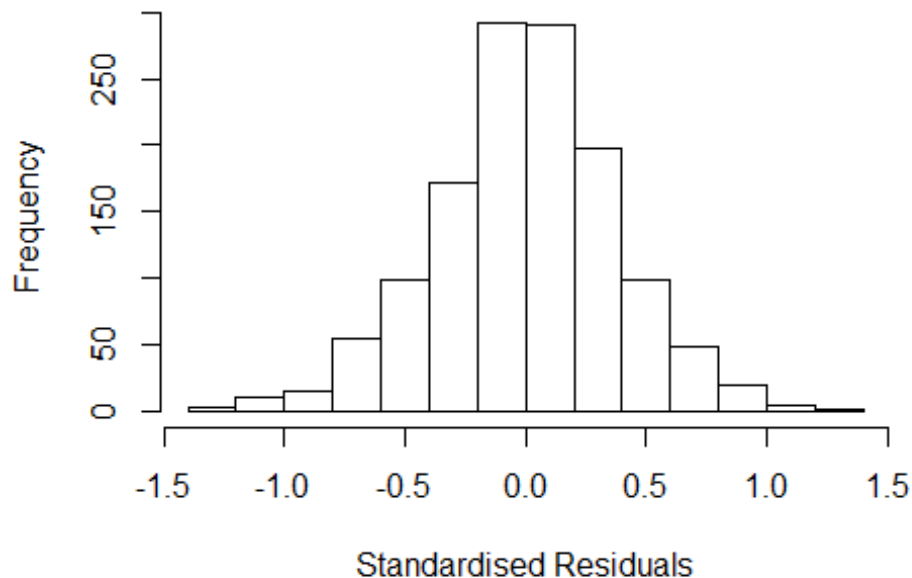


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 6.81887274145525"
## [1] "Male first author team size 2018 geometric mean: 5.83088052929142"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.29209089766647"
## [1] "Male last author team size 2018 geometric mean: 6.59883359442678"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, 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 1.062 1      1.030
## LastAuthorFemale  1.076 1      1.037
## UniqueAuthors     1.391 4      1.042
## Year              1.471 16      1.012
```

Residuals from first and last author and team size



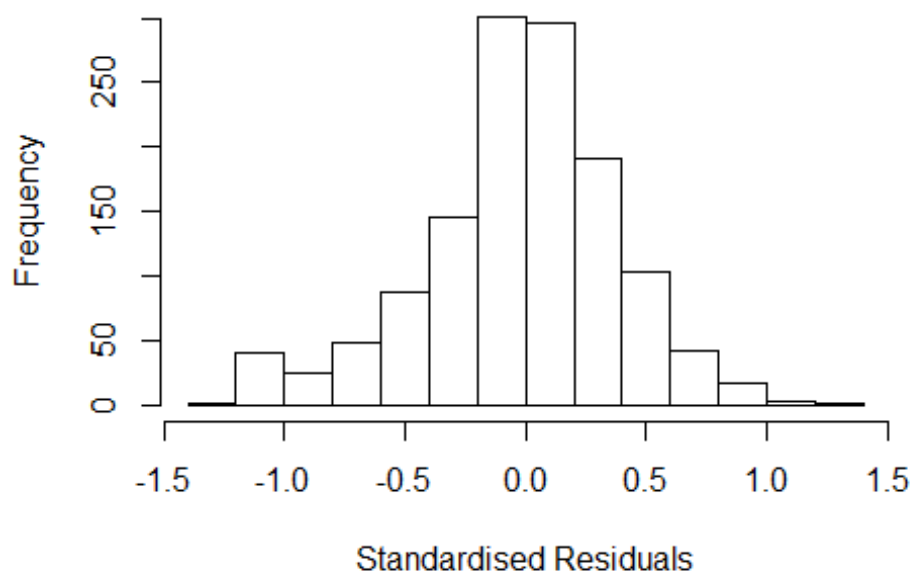
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26071 -0.22907 0.00358 0.23563 1.27924
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.68705 0.09702 7.08 2.3e-12 ***
## FirstAuthorFemale1 -0.06304 0.02076 -3.04 0.0024 **
## LastAuthorFemale1 -0.03369 0.02390 -1.41 0.1588
## UniqueAuthors2 0.48995 0.08721 5.62 2.4e-08 ***
## UniqueAuthors3 0.47241 0.08409 5.62 2.4e-08 ***
## UniqueAuthors4 0.57005 0.08357 6.82 1.4e-11 ***
## UniqueAuthors5 0.66353 0.08099 8.19 6.1e-16 ***
## Year1997 -0.05317 0.07375 -0.72 0.4711
## Year1998 0.06329 0.06596 0.96 0.3374
## Year1999 -0.08252 0.06488 -1.27 0.2037
```

```

## Year2000      0.00361    0.08674    0.04    0.9668
## Year2001     -0.05472    0.07773   -0.70    0.4816
## Year2002     -0.11985    0.06352   -1.89    0.0594 .
## Year2003     -0.12278    0.07553   -1.63    0.1043
## Year2004     -0.09529    0.06454   -1.48    0.1401
## Year2005     -0.03753    0.06200   -0.61    0.5451
## Year2006     -0.14572    0.06372   -2.29    0.0224 *
## Year2007     -0.13540    0.06438   -2.10    0.0356 *
## Year2008     -0.18484    0.06249   -2.96    0.0032 **
## Year2009     -0.11080    0.07051   -1.57    0.1164
## Year2010     -0.14304    0.06285   -2.28    0.0230 *
## Year2011     -0.15546    0.06162   -2.52    0.0118 *
## Year2012     -0.16207    0.06064   -2.67    0.0076 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.191, Adjusted R-squared:  0.177
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1182 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.128  0.857  0.947  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      7.66e-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 1.053 1      1.026
## LastAuthorFemale  1.043 1      1.021
## Year              1.083 16      1.003

```

Residuals from first and last author



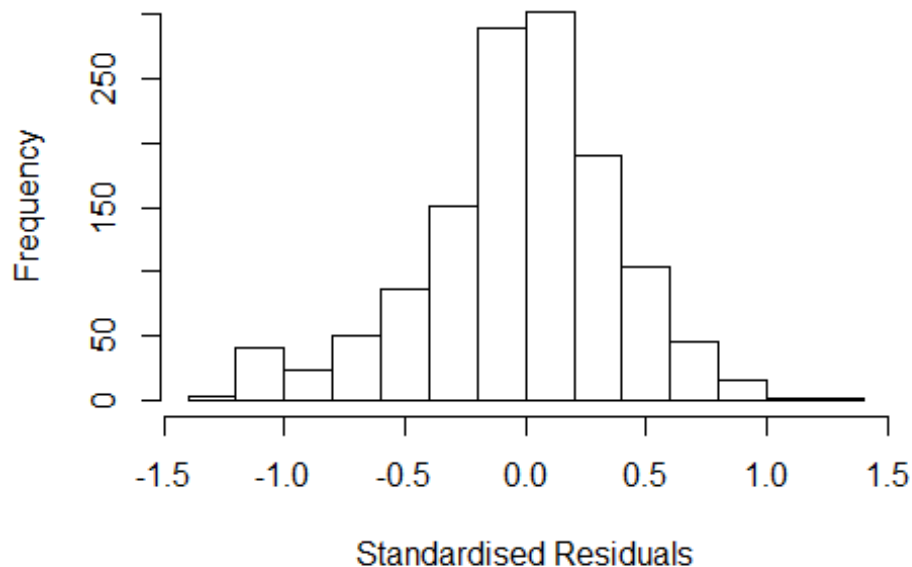
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23084 -0.22416  0.00142  0.22732  1.22332
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.22476    0.05255   23.31  <2e-16 ***
## FirstAuthorFemale1 -0.04460    0.02172   -2.05   0.040 *
## LastAuthorFemale1 -0.04251    0.02591   -1.64   0.101
## Year1997          -0.04522    0.07440   -0.61   0.543
## Year1998           0.05092    0.06543    0.78   0.437
## Year1999          -0.08144    0.06264   -1.30   0.194
## Year2000           0.00608    0.08754    0.07   0.945
## Year2001          -0.06448    0.07769   -0.83   0.407
## Year2002          -0.06709    0.06315   -1.06   0.288
## Year2003          -0.09851    0.07835   -1.26   0.209
## Year2004          -0.03983    0.06729   -0.59   0.554
## Year2005          -0.00418    0.06148   -0.07   0.946
```

```

## Year2006      -0.08994    0.06627   -1.36    0.175
## Year2007      -0.08956    0.06264   -1.43    0.153
## Year2008      -0.16028    0.06341   -2.53    0.012 *
## Year2009      -0.08160    0.07262   -1.12    0.261
## Year2010      -0.06736    0.06439   -1.05    0.296
## Year2011      -0.11892    0.06461   -1.84    0.066 .
## Year2012      -0.10878    0.06280   -1.73    0.083 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0262, Adjusted R-squared:  0.0125
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1176 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.844  0.952  0.879  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      7.66e-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 1.045 1      1.022
## Year      1.045 16      1.001

```

Residuals from first author



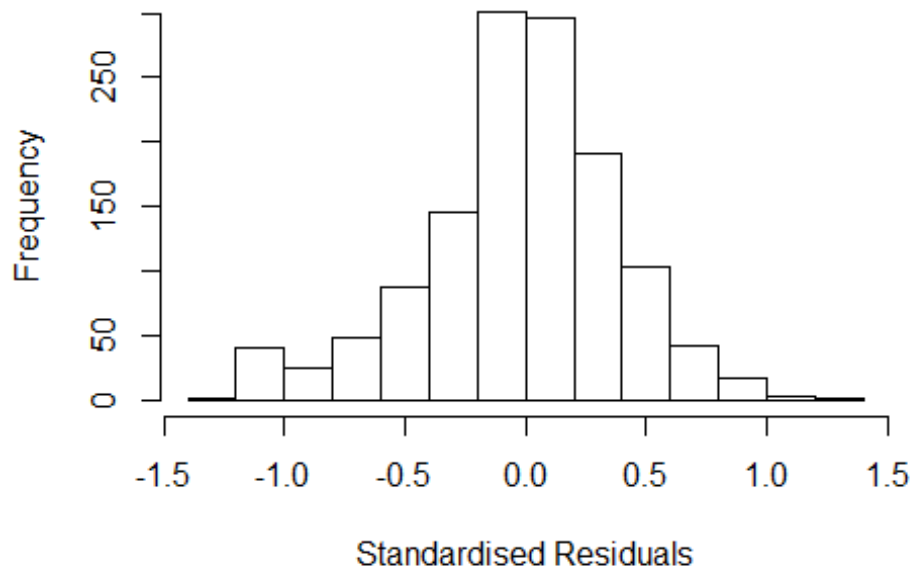
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22720 -0.22720  0.00562  0.22980  1.23010
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.21758    0.05226   23.30  <2e-16 ***
## FirstAuthorFemale1 -0.04622    0.02183   -2.12   0.0345 *
## Year1997         -0.04500    0.07508   -0.60   0.5490
## Year1998          0.05132    0.06563    0.78   0.4344
## Year1999         -0.08198    0.06257   -1.31   0.1904
## Year2000          0.00962    0.08716    0.11   0.9121
## Year2001         -0.06589    0.07810   -0.84   0.3990
## Year2002         -0.06771    0.06338   -1.07   0.2856
## Year2003         -0.09874    0.07884   -1.25   0.2107
## Year2004         -0.04119    0.06738   -0.61   0.5411
## Year2005         -0.00538    0.06157   -0.09   0.9303
## Year2006         -0.09116    0.06619   -1.38   0.1686
```

```

## Year2007          -0.09320    0.06299   -1.48    0.1392
## Year2008          -0.16436    0.06339   -2.59    0.0096 **
## Year2009          -0.08630    0.07248   -1.19    0.2340
## Year2010          -0.06960    0.06466   -1.08    0.2819
## Year2011          -0.12510    0.06498   -1.93    0.0544 .
## Year2012          -0.11342    0.06297   -1.80    0.0719 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.0107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1192 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.190  0.848  0.953  0.880  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      7.66e-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 1.035 1          1.017
## Year              1.035 16          1.001

```


Residuals from last author



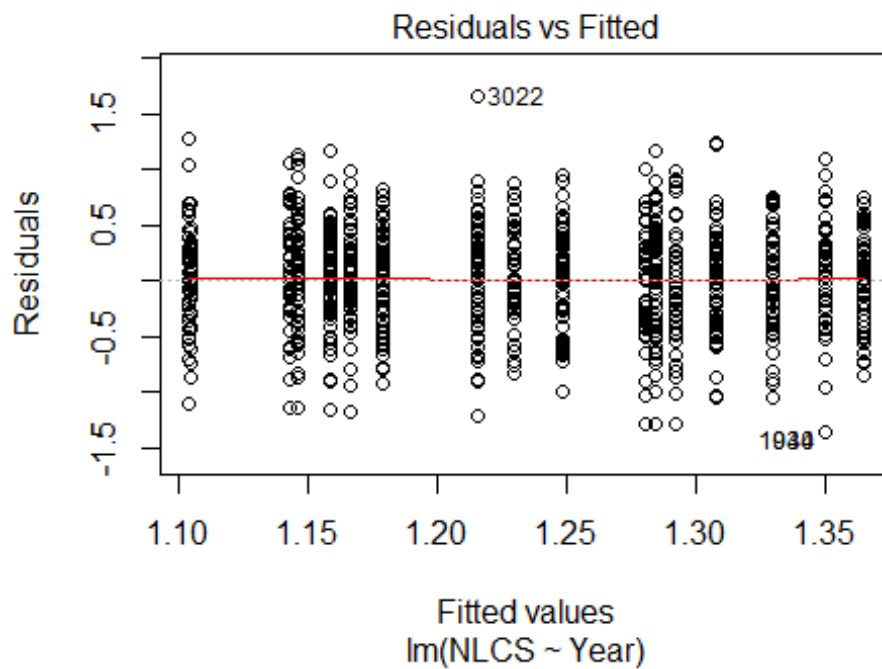
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20966 -0.22734 0.00198 0.23001 1.23876
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20799 0.05204 23.21 <2e-16 ***
## LastAuthorFemale1 -0.04487 0.02595 -1.73 0.0841 .
## Year1997 -0.04485 0.07494 -0.60 0.5496
## Year1998 0.05225 0.06644 0.79 0.4317
## Year1999 -0.08354 0.06396 -1.31 0.1918
## Year2000 0.00167 0.08898 0.02 0.9850
## Year2001 -0.06804 0.07884 -0.86 0.3882
## Year2002 -0.07410 0.06359 -1.17 0.2441
## Year2003 -0.10465 0.08041 -1.30 0.1933
## Year2004 -0.04776 0.06768 -0.71 0.4806
## Year2005 -0.01049 0.06227 -0.17 0.8662
## Year2006 -0.09641 0.06708 -1.44 0.1509
```

```

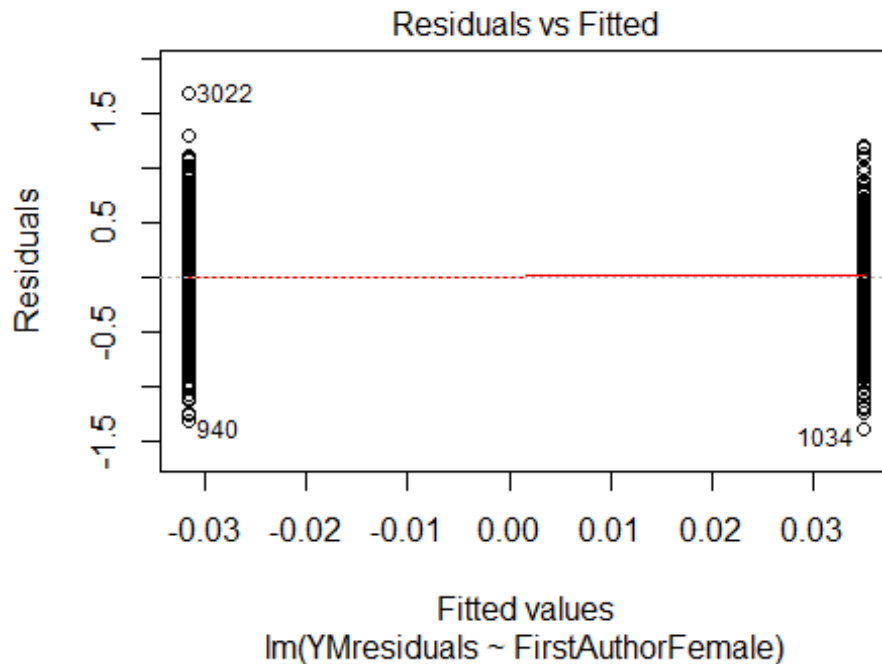
## Year2007          -0.09516      0.06323    -1.51    0.1326
## Year2008          -0.16785      0.06394    -2.63    0.0088 **
## Year2009          -0.08719      0.07294    -1.20    0.2321
## Year2010          -0.07751      0.06529    -1.19    0.2354
## Year2011          -0.12406      0.06510    -1.91    0.0569 .
## Year2012          -0.11754      0.06337    -1.85    0.0639 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.00945
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 135 weights are ~= 1. The remaining 1170 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.185  0.845  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      7.66e-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: 1305"
## [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
## 147 155 119 120 106 135 109 114 137 147 154 149 165 146 172
## 2011 2012
## 167 250
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 37 44 46 54 36 69 48 53 63 91 63 53 66 57 65
## 2011 2012

```

```
## 69 78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 39 36 35 28 49 40 38 52 54 52 47 58 49 55
## 2011 2012
## 58 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 = 26, df = 16, p-value = 0.05
```

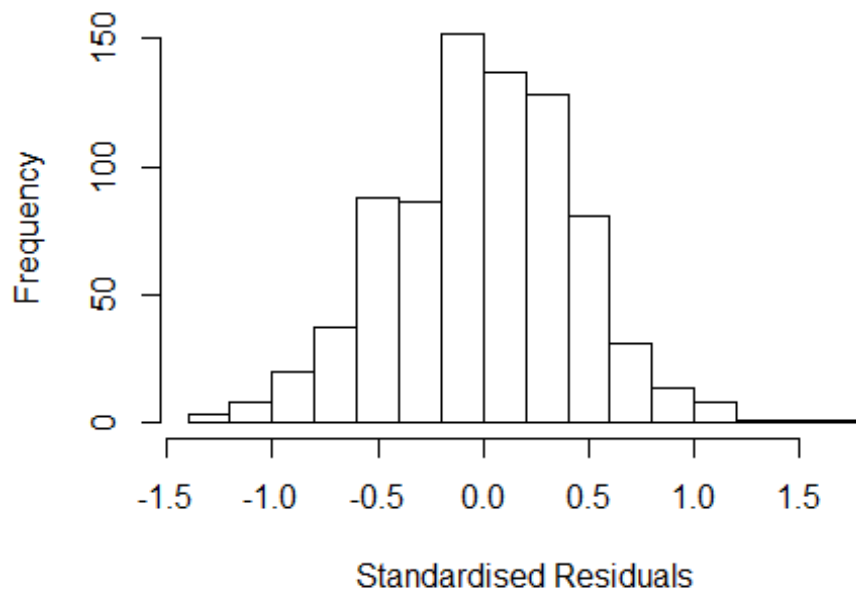


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.1, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 5.58699752970595"
## [1] "Male first author team size 2018 geometric mean: 4.23144734185852"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3900, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.18993993882555"
## [1] "Male last author team size 2018 geometric mean: 4.88610979266312"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, 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 1.161 1      1.078
## LastAuthorFemale  1.109 1      1.053
## UniqueAuthors    1.570 4      1.058
## Year              1.693 16     1.017
```

Residuals from first and last author and team size



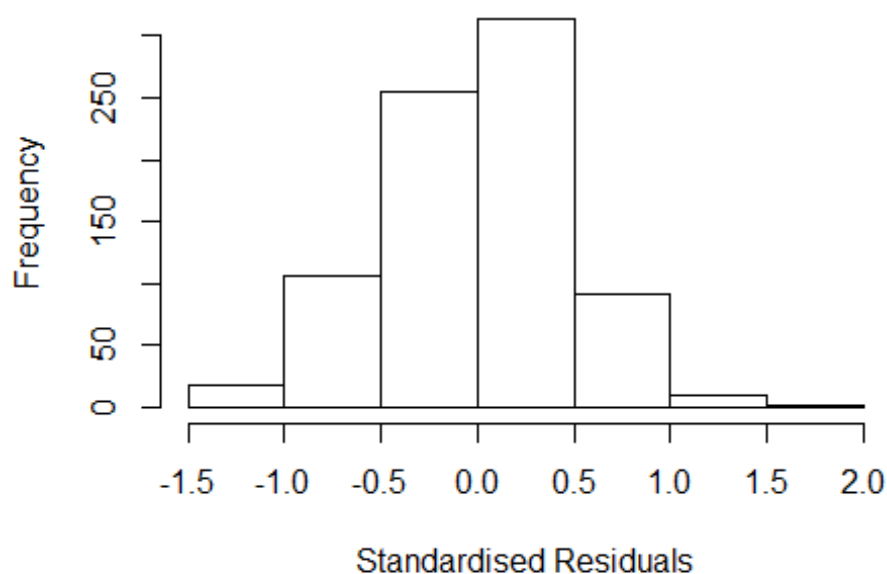
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36247 -0.28469  0.00177  0.28758  1.60925
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.73523    0.09703   7.58 1.0e-13 ***
## FirstAuthorFemale1 0.00265    0.03319   0.08  0.936
## LastAuthorFemale1 -0.06967    0.03541  -1.97  0.049 *
## UniqueAuthors2    0.23554    0.09444   2.49  0.013 *
## UniqueAuthors3    0.45402    0.08265   5.49 5.3e-08 ***
## UniqueAuthors4    0.39131    0.07922   4.94 9.6e-07 ***
## UniqueAuthors5    0.62724    0.07596   8.26 6.5e-16 ***
## Year1997          0.17548    0.12733   1.38  0.169
## Year1998         -0.00908    0.11841  -0.08  0.939
## Year1999          0.21085    0.11485   1.84  0.067 .
```

```

## Year2000      -0.01070    0.10204   -0.10    0.917
## Year2001      0.09471    0.10676    0.89    0.375
## Year2002      0.26721    0.11389    2.35    0.019 *
## Year2003      0.11666    0.12329    0.95    0.344
## Year2004      0.03034    0.10773    0.28    0.778
## Year2005      0.03754    0.10309    0.36    0.716
## Year2006      0.09334    0.09826    0.95    0.342
## Year2007      0.12985    0.10331    1.26    0.209
## Year2008      0.06278    0.10139    0.62    0.536
## Year2009      0.13196    0.10308    1.28    0.201
## Year2010      0.04870    0.10941    0.45    0.656
## Year2011      0.06173    0.11112    0.56    0.579
## Year2012      0.06478    0.10217    0.63    0.526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.18, Adjusted R-squared:  0.157
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.866  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      1.26e-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 1.096 1      1.047
## LastAuthorFemale  1.075 1      1.037
## Year              1.172 16      1.005

```

Residuals from first and last author



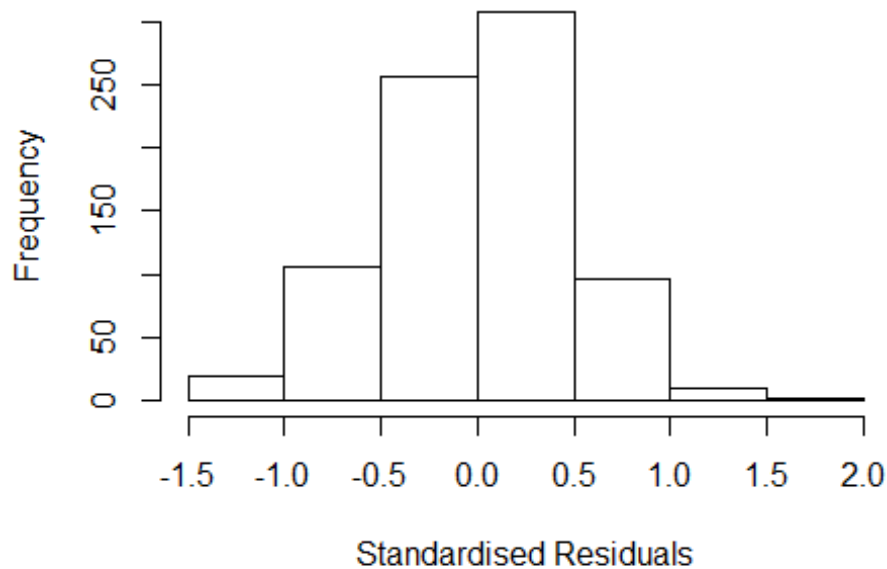
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4601 -0.3407 0.0192 0.3222 1.6474
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0890 0.0994 10.96 <2e-16 ***
## FirstAuthorFemale1 0.0606 0.0351 1.73 0.0848 .
## LastAuthorFemale1 -0.0514 0.0391 -1.31 0.1890
## Year1997 0.2356 0.1427 1.65 0.0991 .
## Year1998 0.0914 0.1326 0.69 0.4910
## Year1999 0.2473 0.1281 1.93 0.0539 .
## Year2000 0.0433 0.1189 0.36 0.7157
## Year2001 0.1587 0.1171 1.36 0.1757
## Year2002 0.3619 0.1221 2.96 0.0031 **
## Year2003 0.2552 0.1290 1.98 0.0483 *
## Year2004 0.1061 0.1175 0.90 0.3670
## Year2005 0.1606 0.1124 1.43 0.1532
```

```

## Year2006          0.1883      0.1161      1.62      0.1052
## Year2007          0.2542      0.1154      2.20      0.0279 *
## Year2008          0.1746      0.1155      1.51      0.1310
## Year2009          0.2841      0.1174      2.42      0.0157 *
## Year2010          0.1921      0.1224      1.57      0.1168
## Year2011          0.0907      0.1319      0.69      0.4919
## Year2012          0.1316      0.1192      1.10      0.2698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.471
## Multiple R-squared:  0.0379, Adjusted R-squared:  0.0156
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 716 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.197  0.867   0.948   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      1.26e-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 1.095 1      1.046
## Year              1.095 16      1.003

```


Residuals from first author



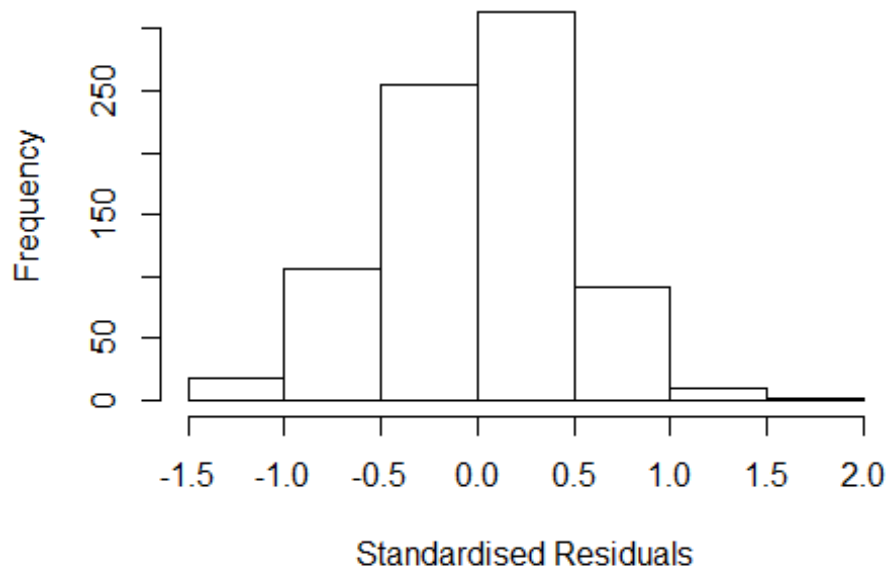
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4998 -0.3326 0.0223 0.3178 1.6690
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0833 0.0981 11.04 <2e-16 ***
## FirstAuthorFemale1 0.0595 0.0352 1.69 0.0914 .
## Year1997 0.2300 0.1407 1.63 0.1025
## Year1998 0.0889 0.1318 0.67 0.5000
## Year1999 0.2429 0.1266 1.92 0.0554 .
## Year2000 0.0316 0.1167 0.27 0.7868
## Year2001 0.1511 0.1159 1.30 0.1926
## Year2002 0.3570 0.1207 2.96 0.0032 **
## Year2003 0.2583 0.1285 2.01 0.0447 *
## Year2004 0.0959 0.1160 0.83 0.4089
## Year2005 0.1569 0.1112 1.41 0.1587
## Year2006 0.1816 0.1150 1.58 0.1147
```

```

## Year2007          0.2452      0.1145      2.14      0.0326 *
## Year2008          0.1674      0.1142      1.47      0.1432
## Year2009          0.2765      0.1159      2.38      0.0173 *
## Year2010          0.1854      0.1211      1.53      0.1261
## Year2011          0.0766      0.1295      0.59      0.5546
## Year2012          0.1157      0.1169      0.99      0.3226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.471
## Multiple R-squared:  0.036, Adjusted R-squared:  0.0149
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 713 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.182  0.868  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      1.26e-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 1.077 1      1.038
## Year              1.077 16      1.002

```

Residuals from last author



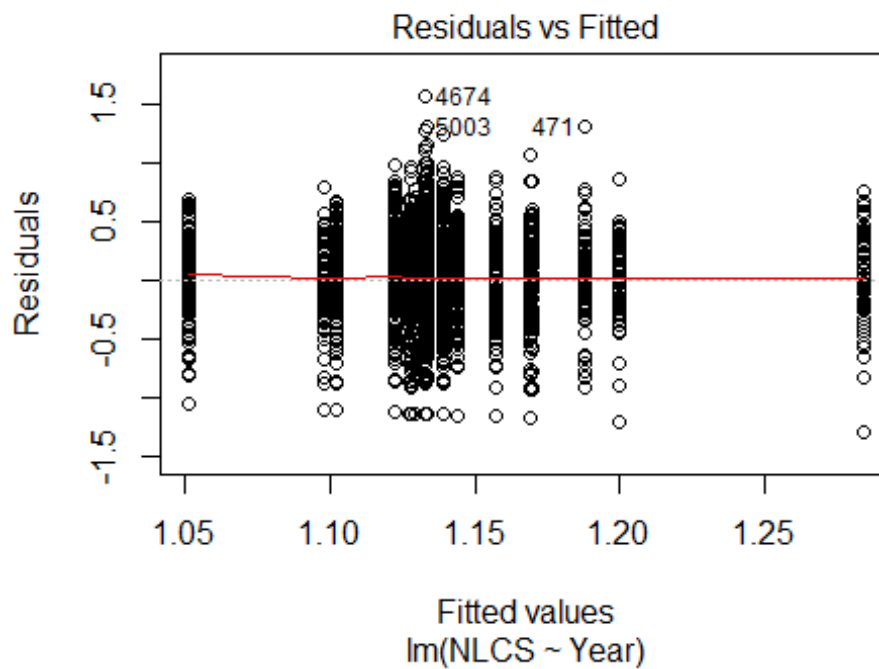
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4370 -0.3344 0.0288 0.3160 1.6129
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1065 0.0992 11.15 <2e-16 ***
## LastAuthorFemale1 -0.0497 0.0390 -1.27 0.2029
## Year1997 0.2391 0.1426 1.68 0.0940 .
## Year1998 0.0954 0.1337 0.71 0.4758
## Year1999 0.2629 0.1272 2.07 0.0391 *
## Year2000 0.0548 0.1201 0.46 0.6481
## Year2001 0.1630 0.1183 1.38 0.1688
## Year2002 0.3801 0.1222 3.11 0.0019 **
## Year2003 0.2651 0.1286 2.06 0.0396 *
## Year2004 0.1193 0.1177 1.01 0.3111
## Year2005 0.1717 0.1133 1.52 0.1300
## Year2006 0.2047 0.1164 1.76 0.0790 .
```

```

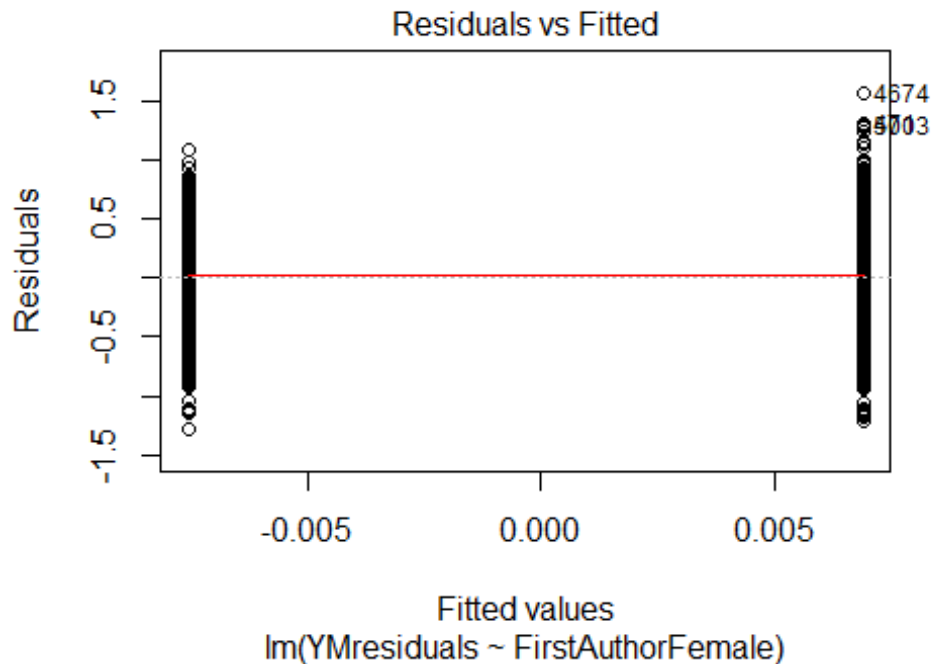
## Year2007          0.2716      0.1157      2.35      0.0192 *
## Year2008          0.1875      0.1159      1.62      0.1060
## Year2009          0.3004      0.1180      2.54      0.0111 *
## Year2010          0.2031      0.1235      1.64      0.1004
## Year2011          0.1084      0.1323      0.82      0.4128
## Year2012          0.1486      0.1191      1.25      0.2126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.0341, Adjusted R-squared:  0.0129
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 714 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.223  0.869  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      1.26e-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: 795"
## [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
##  134  144  158  153  161  163  178  180  181  171  216  227  247  321  331
## 2011 2012
##  365  357
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   56   69   73   78   81   54  117  122  113  121  144  164  164  208  234
## 2011 2012

```

```
## 243 232
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 63 62 69 74 46 105 106 92 106 122 142 142 182 203
## 2011 2012
## 213 210
## [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
```

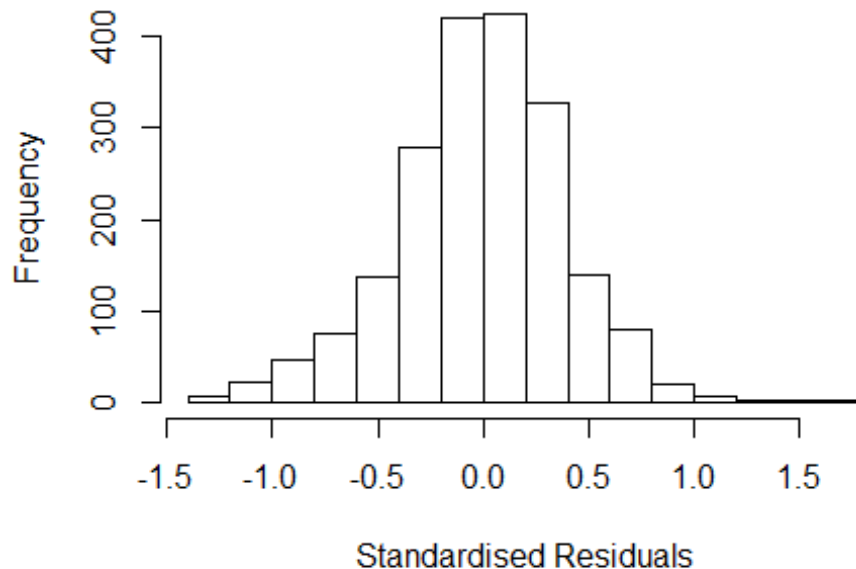


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.7, df = 1, p-value = 0.006
```



```
## [1] "Female first author team size 2018 geometric mean: 5.49989990468631"
## [1] "Male first author team size 2018 geometric mean: 5.07068566071099"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.02542336826762"
## [1] "Male last author team size 2018 geometric mean: 5.4335072858294"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8200, 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 1.059 1          1.029
## LastAuthorFemale  1.106 1          1.052
## UniqueAuthors    1.261 4          1.029
## Year             1.363 16          1.010
```

Residuals from first and last author and team size



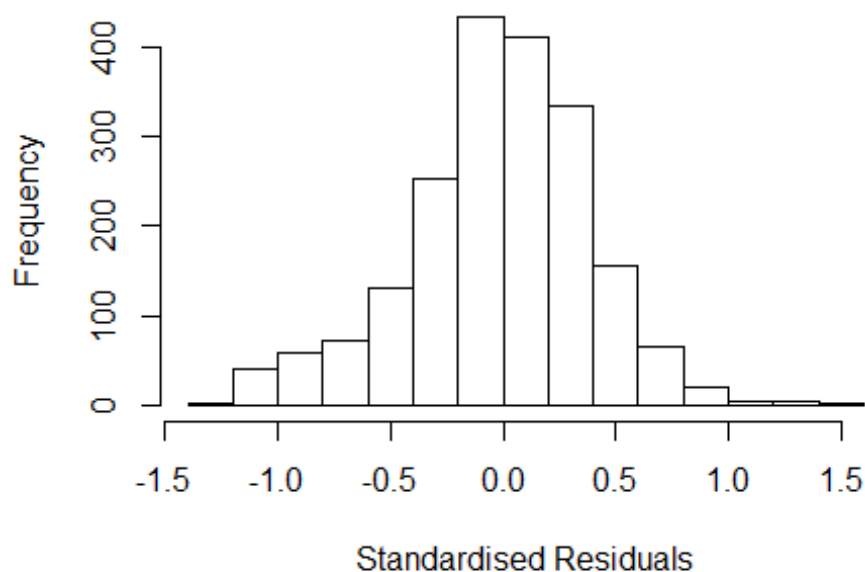
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23566 -0.23869  0.00349  0.24597  1.67518
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1139    0.0777   14.33 < 2e-16 ***
## FirstAuthorFemale1 -0.0103    0.0177   -0.58  0.56209
## LastAuthorFemale1  0.0201    0.0213    0.94  0.34564
## UniqueAuthors2     0.1233    0.0584    2.11  0.03496 *
## UniqueAuthors3     0.1436    0.0557    2.58  0.00993 **
## UniqueAuthors4     0.2372    0.0553    4.29  1.9e-05 ***
## UniqueAuthors5     0.2915    0.0522    5.58  2.7e-08 ***
## Year1997          -0.0664    0.0710   -0.94  0.34972
## Year1998          -0.1142    0.0769   -1.48  0.13777
## Year1999          -0.1997    0.0711   -2.81  0.00502 **
```

```

## Year2000          -0.1594      0.0792    -2.01  0.04443 *
## Year2001          -0.1623      0.0888    -1.83  0.06779 .
## Year2002          -0.1684      0.0698    -2.41  0.01596 *
## Year2003          -0.2480      0.0699    -3.55  0.00040 ***
## Year2004          -0.1808      0.0696    -2.60  0.00940 **
## Year2005          -0.2601      0.0717    -3.63  0.00029 ***
## Year2006          -0.1697      0.0686    -2.47  0.01352 *
## Year2007          -0.1826      0.0664    -2.75  0.00599 **
## Year2008          -0.2045      0.0664    -3.08  0.00209 **
## Year2009          -0.2110      0.0679    -3.11  0.00192 **
## Year2010          -0.2120      0.0695    -3.05  0.00232 **
## Year2011          -0.2307      0.0687    -3.36  0.00080 ***
## Year2012          -0.1910      0.0679    -2.81  0.00499 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0594, Adjusted R-squared:  0.0488
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 164 weights are ~= 1. The remaining 1824 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8650 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          5.03e-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 1.045 1 1.022
## LastAuthorFemale 1.104 1 1.051
## Year 1.139 16 1.004

```


Residuals from first and last author



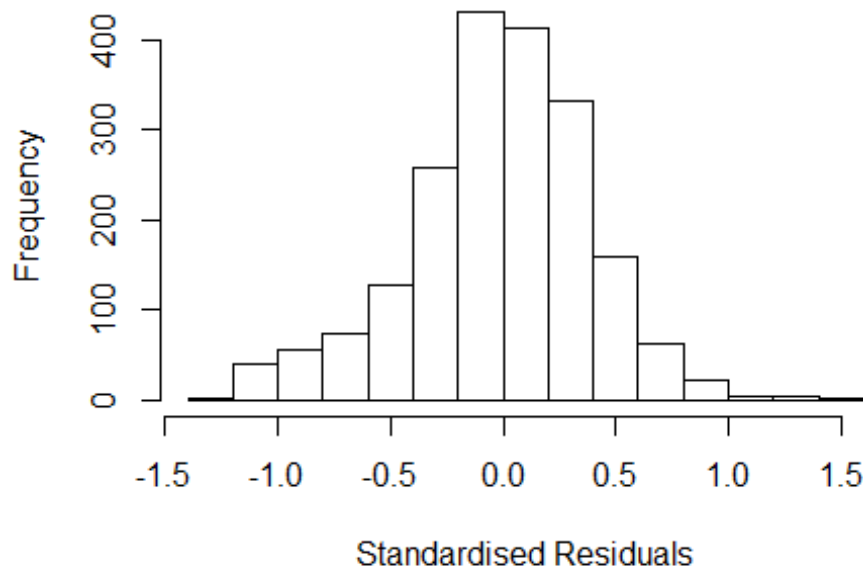
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23819 -0.23352  0.00305  0.24255  1.57005
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.29749    0.05723   22.67  <2e-16 ***
## FirstAuthorFemale1 -0.00812    0.01782   -0.46   0.6488
## LastAuthorFemale1  0.01006    0.02159    0.47   0.6413
## Year1997        -0.05930    0.06808   -0.87   0.3838
## Year1998        -0.11058    0.07339   -1.51   0.1320
## Year1999        -0.18774    0.06786   -2.77   0.0057 **
## Year2000        -0.13102    0.07769   -1.69   0.0919 .
## Year2001        -0.11162    0.08864   -1.26   0.2081
## Year2002        -0.12158    0.06680   -1.82   0.0689 .
## Year2003        -0.19700    0.06636   -2.97   0.0030 **
## Year2004        -0.13015    0.06688   -1.95   0.0518 .
## Year2005        -0.22493    0.06851   -3.28   0.0010 **
```

```

## Year2006      -0.11892    0.06535   -1.82    0.0689 .
## Year2007      -0.14604    0.06274   -2.33    0.0200 *
## Year2008      -0.16164    0.06259   -2.58    0.0099 **
## Year2009      -0.15512    0.06452   -2.40    0.0163 *
## Year2010      -0.15240    0.06637   -2.30    0.0218 *
## Year2011      -0.16555    0.06491   -2.55    0.0108 *
## Year2012      -0.14260    0.06437   -2.22    0.0268 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.363
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00178
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 171 weights are ~= 1. The remaining 1817 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0222 0.8620 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      5.03e-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 1.039 1      1.019
## Year      1.039 16      1.001

```

Residuals from first author



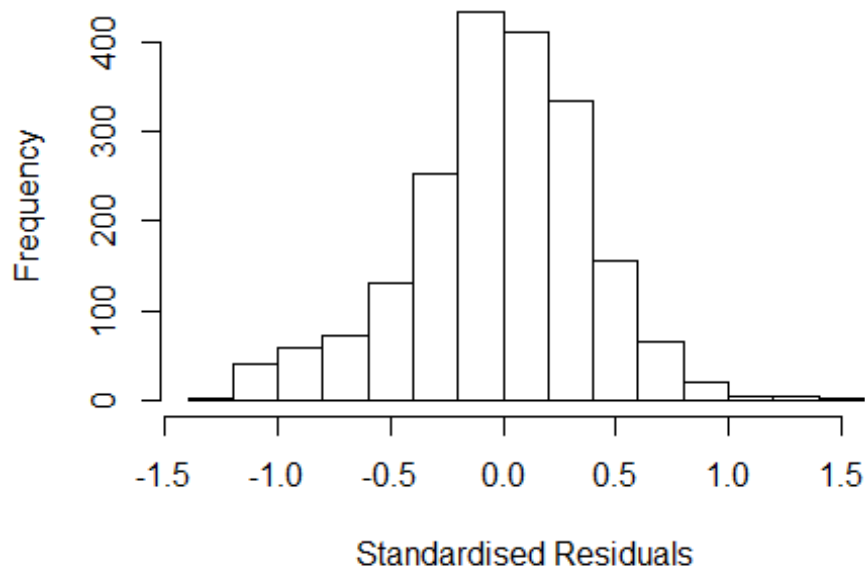
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23908 -0.23472 0.00466 0.24283 1.56763
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29780 0.05716 22.71 <2e-16 ***
## FirstAuthorFemale1 -0.00712 0.01780 -0.40 0.6893
## Year1997 -0.05871 0.06802 -0.86 0.3882
## Year1998 -0.10841 0.07322 -1.48 0.1389
## Year1999 -0.18643 0.06781 -2.75 0.0060 **
## Year2000 -0.13092 0.07769 -1.69 0.0921 .
## Year2001 -0.10968 0.08833 -1.24 0.2145
## Year2002 -0.11994 0.06674 -1.80 0.0725 .
## Year2003 -0.19522 0.06613 -2.95 0.0032 **
## Year2004 -0.12895 0.06689 -1.93 0.0540 .
## Year2005 -0.22348 0.06846 -3.26 0.0011 **
## Year2006 -0.11787 0.06526 -1.81 0.0710 .
```

```

## Year2007          -0.14409      0.06263    -2.30    0.0215 *
## Year2008          -0.15997      0.06249    -2.56    0.0105 *
## Year2009          -0.15271      0.06424    -2.38    0.0175 *
## Year2010          -0.15109      0.06639    -2.28    0.0230 *
## Year2011          -0.16343      0.06479    -2.52    0.0117 *
## Year2012          -0.13978      0.06388    -2.19    0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.364
## Multiple R-squared:  0.0107, Adjusted R-squared:  0.00216
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1814 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0233 0.8630 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      5.03e-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 1.097 1          1.047
## Year            1.097 16          1.003

```

Residuals from last author



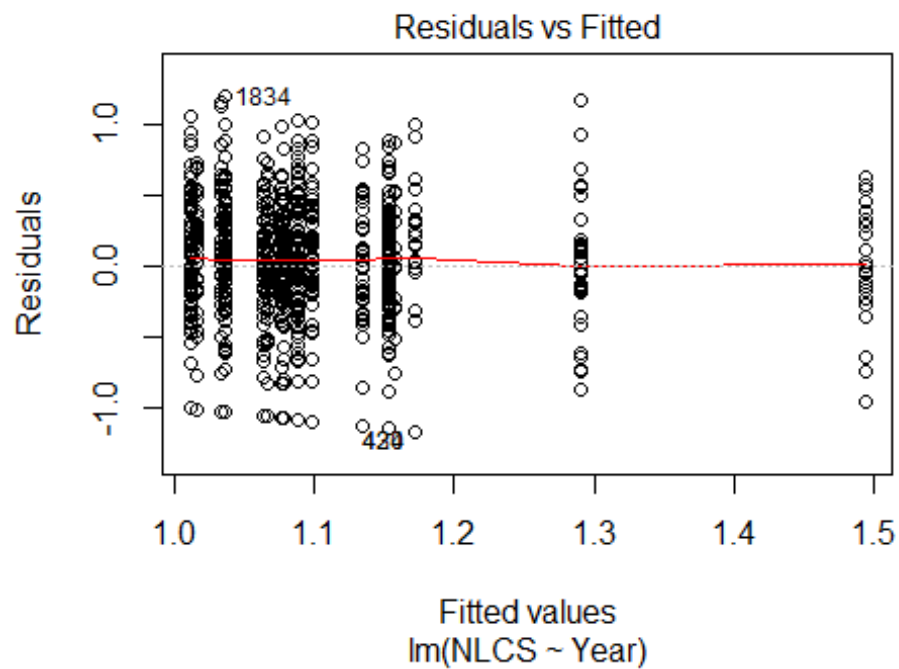
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23549 -0.23330 0.00301 0.24350 1.57433
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29320 0.05650 22.89 <2e-16 ***
## LastAuthorFemale1 0.00899 0.02155 0.42 0.6765
## Year1997 -0.05771 0.06799 -0.85 0.3961
## Year1998 -0.10934 0.07319 -1.49 0.1354
## Year1999 -0.18577 0.06766 -2.75 0.0061 **
## Year2000 -0.12953 0.07766 -1.67 0.0955 .
## Year2001 -0.11000 0.08864 -1.24 0.2148
## Year2002 -0.12074 0.06678 -1.81 0.0708 .
## Year2003 -0.19603 0.06636 -2.95 0.0032 **
## Year2004 -0.12969 0.06683 -1.94 0.0524 .
## Year2005 -0.22426 0.06846 -3.28 0.0011 **
## Year2006 -0.11826 0.06532 -1.81 0.0704 .
```

```

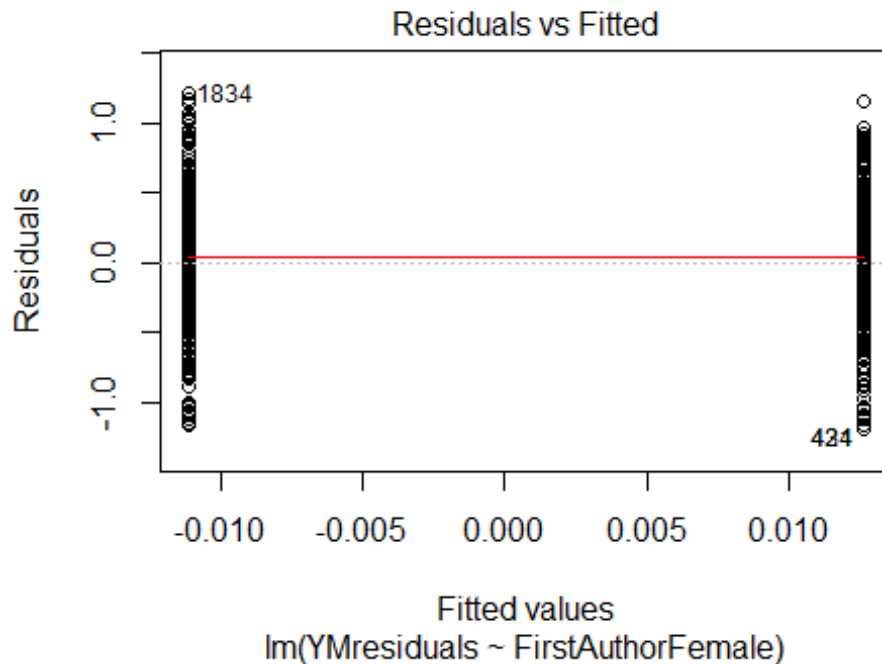
## Year2007          -0.14587      0.06263    -2.33    0.0199 *
## Year2008          -0.16090      0.06255    -2.57    0.0102 *
## Year2009          -0.15429      0.06443    -2.39    0.0167 *
## Year2010          -0.15211      0.06631    -2.29    0.0219 *
## Year2011          -0.16553      0.06481    -2.55    0.0107 *
## Year2012          -0.14249      0.06426    -2.22    0.0267 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.363
## Multiple R-squared:  0.0107, Adjusted R-squared:  0.00221
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 175 weights are ~= 1. The remaining 1813 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0205 0.8610 0.9520 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      5.03e-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: 1988"
## [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
##   75  106  93   73   98   70   89   74   73   72   96   87   94  107  133
## 2011 2012
##  134  123
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   38   35   41   24   16   47   50   53   45   60   58   62   65   89
## 2011 2012

```

```
## 91 84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 32 32 39 23 14 41 40 46 43 51 46 54 56 79
## 2011 2012
## 83 77
## [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
```

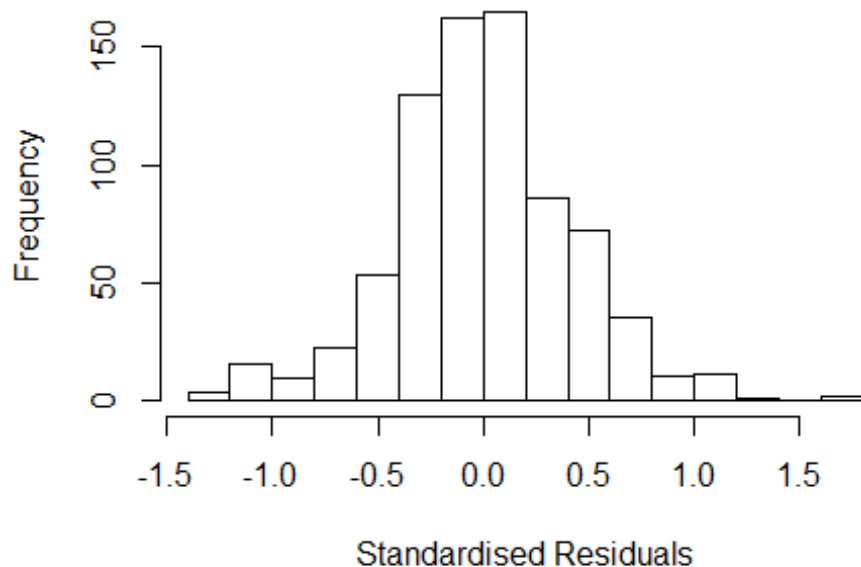


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.93999908917739"
## [1] "Male first author team size 2018 geometric mean: 4.40777416041617"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.11151527839492"
## [1] "Male last author team size 2018 geometric mean: 4.47972703504249"
##
## 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 1.262 1      1.123
## LastAuthorFemale  1.140 1      1.068
## UniqueAuthors     1.689 4      1.068
## Year              1.916 16      1.021
```


Residuals from first and last author and team size



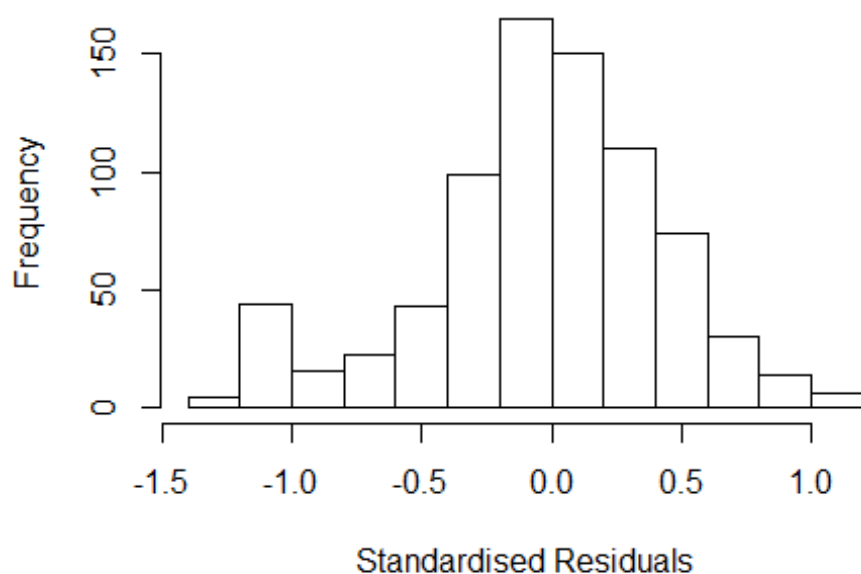
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29233 -0.24110 -0.00825  0.24958  1.72637
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.75619    0.10346   7.31 6.9e-13 ***
## FirstAuthorFemale1  0.01625    0.03040   0.53 0.59316
## LastAuthorFemale1 -0.00683    0.03278  -0.21 0.83512
## UniqueAuthors2     0.73611    0.09316   7.90 9.8e-15 ***
## UniqueAuthors3     0.71358    0.08558   8.34 3.6e-16 ***
## UniqueAuthors4     0.82350    0.08215  10.02 < 2e-16 ***
## UniqueAuthors5     0.88274    0.07787  11.34 < 2e-16 ***
## Year1997          -0.42811    0.12070  -3.55 0.00041 ***
## Year1998          -0.23871    0.09687  -2.46 0.01396 *
## Year1999          -0.26113    0.10304  -2.53 0.01147 *
```

```

## Year2000      -0.28736      0.12755      -2.25      0.02455 *
## Year2001      -0.39662      0.15588      -2.54      0.01114 *
## Year2002      -0.43798      0.09240      -4.74      2.6e-06 ***
## Year2003      -0.46267      0.10062      -4.60      5.0e-06 ***
## Year2004      -0.44886      0.08735      -5.14      3.5e-07 ***
## Year2005      -0.46198      0.09278      -4.98      7.9e-07 ***
## Year2006      -0.39988      0.08580      -4.66      3.7e-06 ***
## Year2007      -0.46500      0.08793      -5.29      1.6e-07 ***
## Year2008      -0.47002      0.08517      -5.52      4.7e-08 ***
## Year2009      -0.38941      0.09648      -4.04      6.0e-05 ***
## Year2010      -0.45939      0.09436      -4.87      1.4e-06 ***
## Year2011      -0.47542      0.08452      -5.62      2.6e-08 ***
## Year2012      -0.53447      0.08719      -6.13      1.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.294, Adjusted R-squared:  0.274
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 177 is an outlier with |weight| = 0 ( < 0.00013);
## 63 weights are ~= 1. The remaining 712 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0029 0.8640 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           1.29e-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 1.192 1 1.092
## LastAuthorFemale 1.067 1 1.033
## Year 1.211 16 1.006

```

Residuals from first and last author



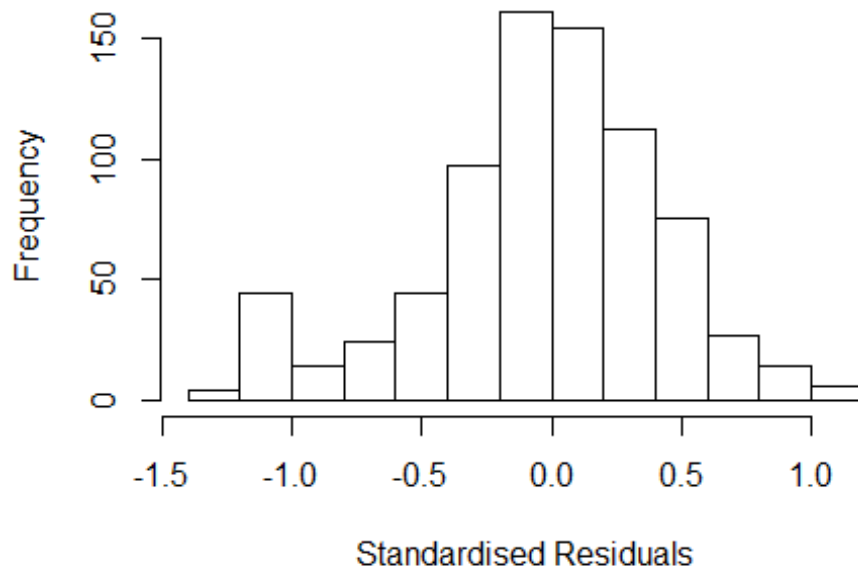
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33719 -0.25221 -0.00231  0.25630  1.17372
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5323     0.0868   17.65 < 2e-16 ***
## FirstAuthorFemale1  0.0363     0.0330    1.10  0.27216
## LastAuthorFemale1 -0.0114     0.0381   -0.30  0.76580
## Year1997          -0.4396     0.1376   -3.19  0.00146 **
## Year1998          -0.2499     0.1151   -2.17  0.03030 *
## Year1999          -0.3760     0.1045   -3.60  0.00034 ***
## Year2000          -0.2201     0.1460   -1.51  0.13227
## Year2001          -0.3906     0.1630   -2.40  0.01680 *
## Year2002          -0.4289     0.1066   -4.02  6.3e-05 ***
## Year2003          -0.4378     0.1050   -4.17  3.4e-05 ***
## Year2004          -0.4067     0.1005   -4.04  5.8e-05 ***
## Year2005          -0.4836     0.1080   -4.48  8.7e-06 ***
```

```

## Year2006          -0.3674      0.0986   -3.73  0.00021 ***
## Year2007          -0.4765      0.1046   -4.56  6.1e-06 ***
## Year2008          -0.4519      0.0964   -4.69  3.3e-06 ***
## Year2009          -0.3791      0.1095   -3.46  0.00057 ***
## Year2010          -0.4525      0.1071   -4.22  2.7e-05 ***
## Year2011          -0.4720      0.0989   -4.77  2.2e-06 ***
## Year2012          -0.5256      0.1053   -4.99  7.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0497, Adjusted R-squared:  0.0271
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 711 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.248  0.858  0.956  0.884  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      1.29e-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 1.159 1      1.077
## Year              1.159 16      1.005

```

Residuals from first author



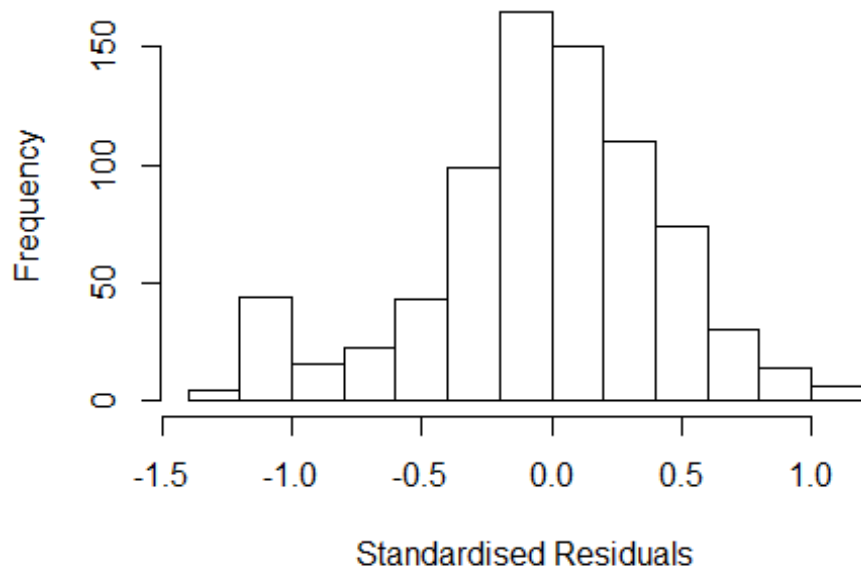
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35e+00 -2.49e-01  1.52e-05  2.54e-01  1.18e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5306     0.0862   17.75 < 2e-16 ***
## FirstAuthorFemale1  0.0363     0.0330    1.10  0.27296
## Year1997        -0.4406     0.1375   -3.20  0.00141 **
## Year1998        -0.2517     0.1150   -2.19  0.02900 *
## Year1999        -0.3760     0.1044   -3.60  0.00034 ***
## Year2000        -0.2207     0.1454   -1.52  0.12951
## Year2001        -0.3925     0.1630   -2.41  0.01628 *
## Year2002        -0.4302     0.1064   -4.04  5.8e-05 ***
## Year2003        -0.4386     0.1047   -4.19  3.2e-05 ***
## Year2004        -0.4071     0.1003   -4.06  5.4e-05 ***
## Year2005        -0.4840     0.1078   -4.49  8.2e-06 ***
## Year2006        -0.3680     0.0983   -3.74  0.00019 ***
```

```

## Year2007          -0.4766      0.1044   -4.57  5.8e-06 ***
## Year2008          -0.4531      0.0961   -4.72  2.9e-06 ***
## Year2009          -0.3808      0.1094   -3.48  0.00053 ***
## Year2010          -0.4551      0.1064   -4.28  2.1e-05 ***
## Year2011          -0.4737      0.0987   -4.80  1.9e-06 ***
## Year2012          -0.5275      0.1050   -5.02  6.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0497, Adjusted R-squared:  0.0284
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 705 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.239  0.858  0.954  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      1.29e-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 1.039 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
## -1.32302 -0.25201 -0.00615 0.25528 1.15487
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5499 0.0853 18.17 < 2e-16 ***
## LastAuthorFemale1 -0.0111 0.0382 -0.29 0.77135
## Year1997 -0.4403 0.1378 -3.19 0.00146 **
## Year1998 -0.2490 0.1145 -2.17 0.02995 *
## Year1999 -0.3808 0.1046 -3.64 0.00029 ***
## Year2000 -0.2269 0.1429 -1.59 0.11289
## Year2001 -0.3968 0.1634 -2.43 0.01542 *
## Year2002 -0.4304 0.1062 -4.05 5.6e-05 ***
## Year2003 -0.4409 0.1047 -4.21 2.9e-05 ***
## Year2004 -0.4070 0.1001 -4.06 5.3e-05 ***
## Year2005 -0.4782 0.1078 -4.44 1.1e-05 ***
## Year2006 -0.3711 0.0983 -3.78 0.00017 ***
```

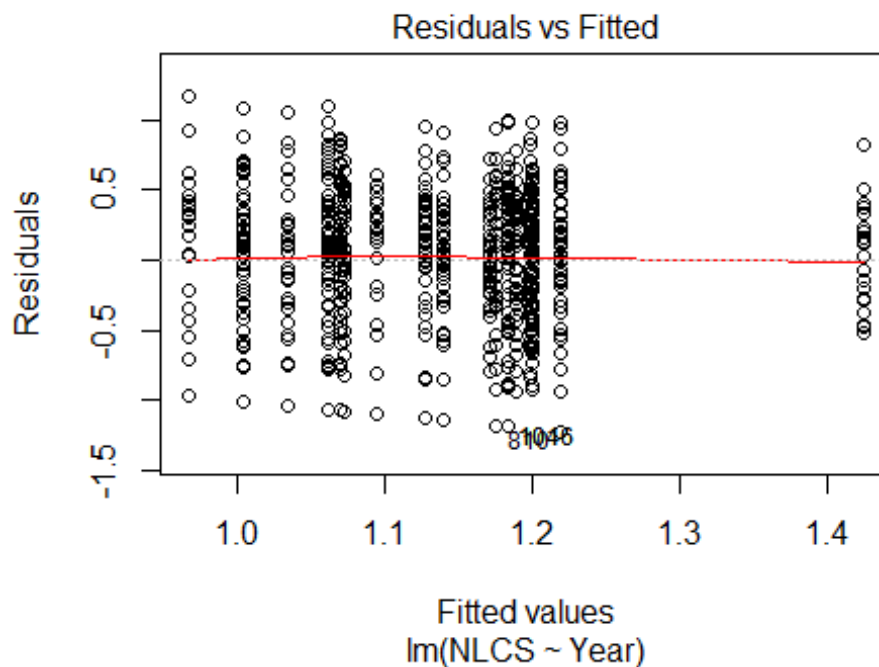
```

## Year2007          -0.4781      0.1052    -4.54  6.5e-06 ***
## Year2008          -0.4536      0.0963    -4.71  3.0e-06 ***
## Year2009          -0.3778      0.1098    -3.44  0.00061 ***
## Year2010          -0.4441      0.1061    -4.19  3.2e-05 ***
## Year2011          -0.4708      0.0987    -4.77  2.2e-06 ***
## Year2012          -0.5229      0.1055    -4.96  8.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0478, Adjusted R-squared:  0.0264
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 711 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.256  0.860  0.956  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      1.29e-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: 776"
## [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
##   54   40   43   33   44   80   55   54   65   74   64   72   58   90   81
## 2011 2012
##   95   95
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   37   22   28   23   27   46   39   42   50   60   46   53   47   70   62
## 2011 2012

```



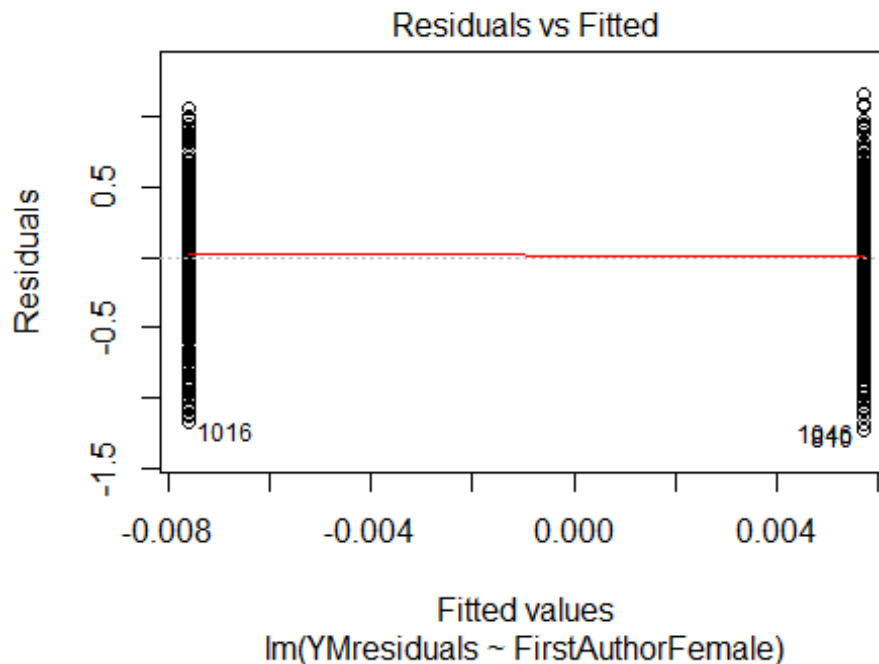
```
## 79 74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 32 16 24 21 26 44 34 40 48 54 44 47 45 62 57
## 2011 2012
## 72 66
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, 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: 4.17849837479394"
## [1] "Male first author team size 2018 geometric mean: 4.72322589545786"
## 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 = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.58805532033149"
## [1] "Male last author team size 2018 geometric mean: 4.40175451880919"

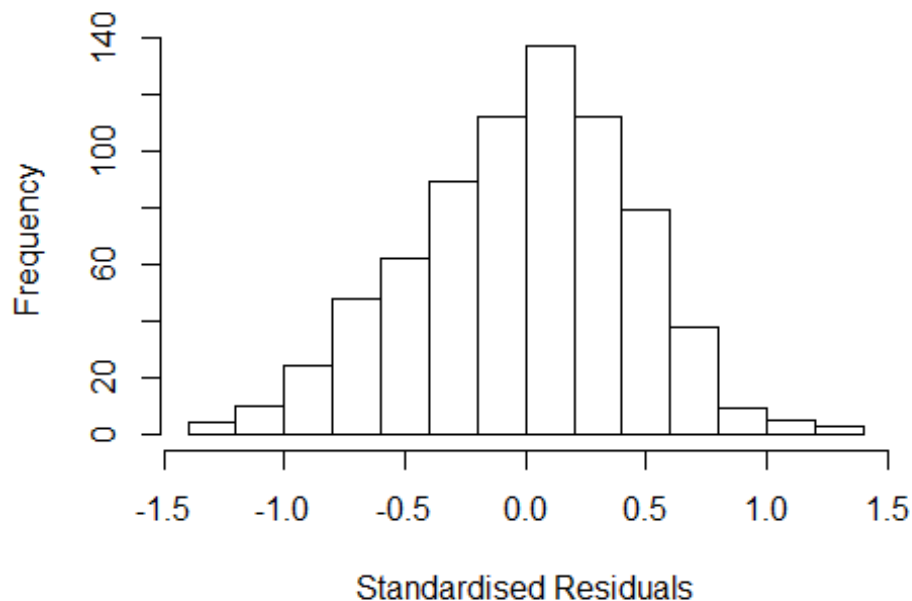
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 420, p-value = 0.6
## 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	1.146	1	1.070
LastAuthorFemale	1.130	1	1.063
UniqueAuthors	1.646	4	1.064
Year	1.703	16	1.017

Residuals from first and last author and team size



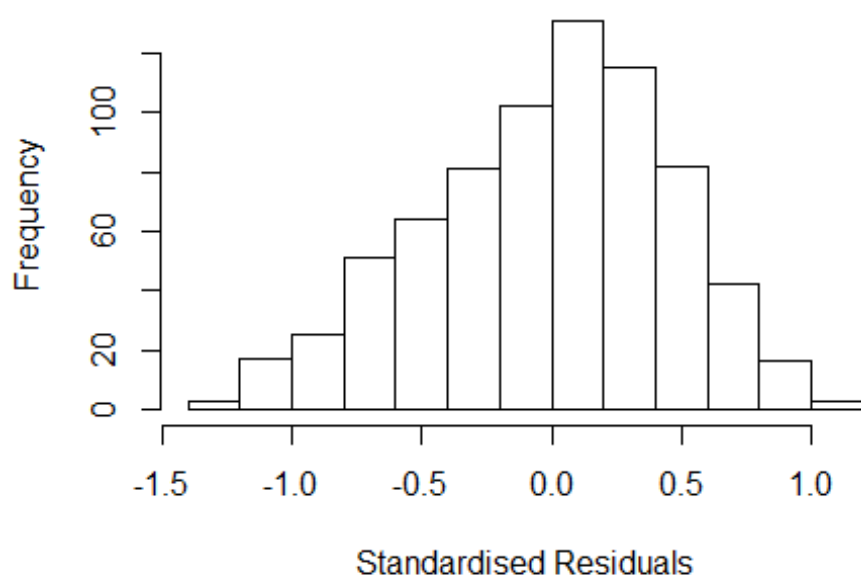
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3088 -0.3032 0.0207 0.3023 1.3452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86657 0.12706 6.82 2.0e-11 ***
## FirstAuthorFemale1 -0.04002 0.03695 -1.08 0.2791
## LastAuthorFemale1 -0.03813 0.04280 -0.89 0.3732
## UniqueAuthors2 0.18610 0.10033 1.85 0.0640 .
## UniqueAuthors3 0.25343 0.08738 2.90 0.0038 **
## UniqueAuthors4 0.25933 0.08845 2.93 0.0035 **
## UniqueAuthors5 0.44219 0.07681 5.76 1.3e-08 ***
## Year1997 -0.08242 0.15775 -0.52 0.6015
## Year1998 -0.22878 0.16247 -1.41 0.1595
## Year1999 0.28104 0.12841 2.19 0.0290 *
```

```

## Year2000      0.03463    0.14068    0.25    0.8057
## Year2001      0.04919    0.12642    0.39    0.6973
## Year2002     -0.06191    0.12937   -0.48    0.6324
## Year2003     -0.02477    0.11455   -0.22    0.8289
## Year2004      0.00840    0.11674    0.07    0.9426
## Year2005     -0.15066    0.12181   -1.24    0.2166
## Year2006     -0.03392    0.12042   -0.28    0.7783
## Year2007      0.07553    0.12445    0.61    0.5441
## Year2008      0.00294    0.11999    0.02    0.9804
## Year2009      0.01050    0.11885    0.09    0.9296
## Year2010      0.08108    0.12429    0.65    0.5144
## Year2011     -0.04227    0.11931   -0.35    0.7232
## Year2012     -0.08885    0.11817   -0.75    0.4523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0799
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 671 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.362  0.873   0.954   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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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  1.061 1      1.030
## Year              1.154 16      1.004

```

Residuals from first and last author



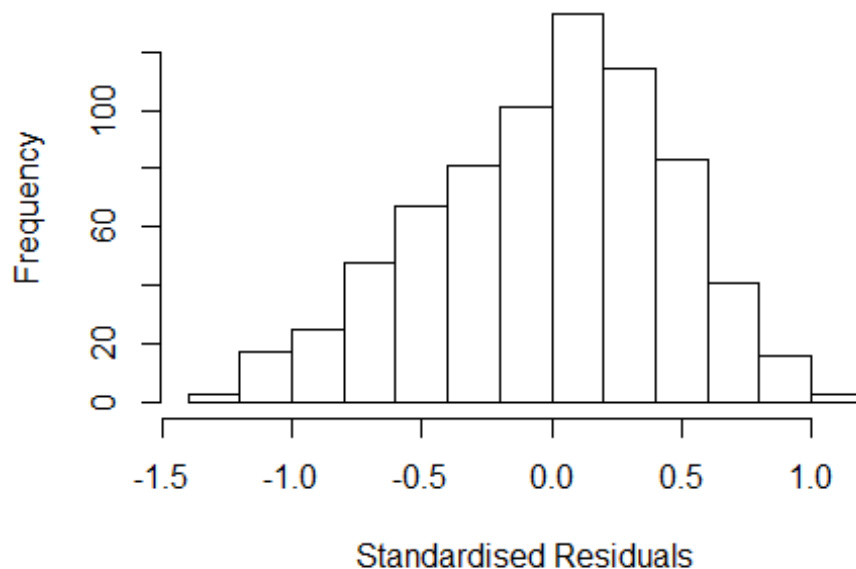
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.256 -0.329  0.033  0.320  1.176
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11747    0.09329   11.98  <2e-16 ***
## FirstAuthorFemale1 -0.02050    0.03824   -0.54    0.592
## LastAuthorFemale1 -0.00863    0.04509   -0.19    0.848
## Year1997        -0.01356    0.15773   -0.09    0.932
## Year1998        -0.15408    0.15295   -1.01    0.314
## Year1999         0.33717    0.11964    2.82    0.005 **
## Year2000         0.11878    0.13785    0.86    0.389
## Year2001         0.10284    0.12024    0.86    0.393
## Year2002         0.01598    0.12519    0.13    0.898
## Year2003         0.05255    0.10690    0.49    0.623
## Year2004         0.08632    0.11212    0.77    0.442
## Year2005        -0.08146    0.11675   -0.70    0.486
```

```

## Year2006          0.05602    0.11616    0.48    0.630
## Year2007          0.13526    0.12003    1.13    0.260
## Year2008          0.07275    0.11625    0.63    0.532
## Year2009          0.10724    0.11048    0.97    0.332
## Year2010          0.13882    0.11968    1.16    0.246
## Year2011         -0.03593    0.11667   -0.31    0.758
## Year2012         -0.00691    0.11301   -0.06    0.951
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.477
## Multiple R-squared:  0.0348, Adjusted R-squared:  0.0105
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 667 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.468  0.871  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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0        1000         0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



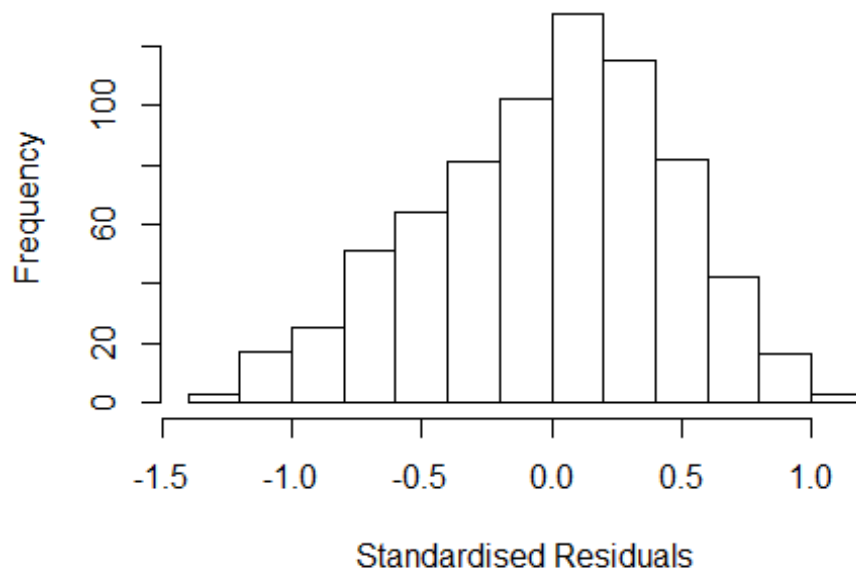
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2547 -0.3315 0.0278 0.3189 1.1692
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11642 0.09341 11.95 <2e-16 ***
## FirstAuthorFemale1 -0.02100 0.03833 -0.55 0.584
## Year1997 -0.01427 0.15760 -0.09 0.928
## Year1998 -0.15462 0.15307 -1.01 0.313
## Year1999 0.33679 0.11973 2.81 0.005 **
## Year2000 0.11816 0.13753 0.86 0.391
## Year2001 0.10205 0.11971 0.85 0.394
## Year2002 0.01571 0.12513 0.13 0.900
## Year2003 0.05226 0.10681 0.49 0.625
## Year2004 0.08627 0.11210 0.77 0.442
## Year2005 -0.08264 0.11619 -0.71 0.477
## Year2006 0.05569 0.11612 0.48 0.632
```

```

## Year2007          0.13447    0.11965    1.12    0.261
## Year2008          0.07240    0.11626    0.62    0.534
## Year2009          0.10636    0.11018    0.97    0.335
## Year2010          0.13833    0.11942    1.16    0.247
## Year2011         -0.03761    0.11617   -0.32    0.746
## Year2012         -0.00859    0.11199   -0.08    0.939
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.476
## Multiple R-squared:  0.0348, Adjusted R-squared:  0.0118
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 668 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.467  0.870   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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



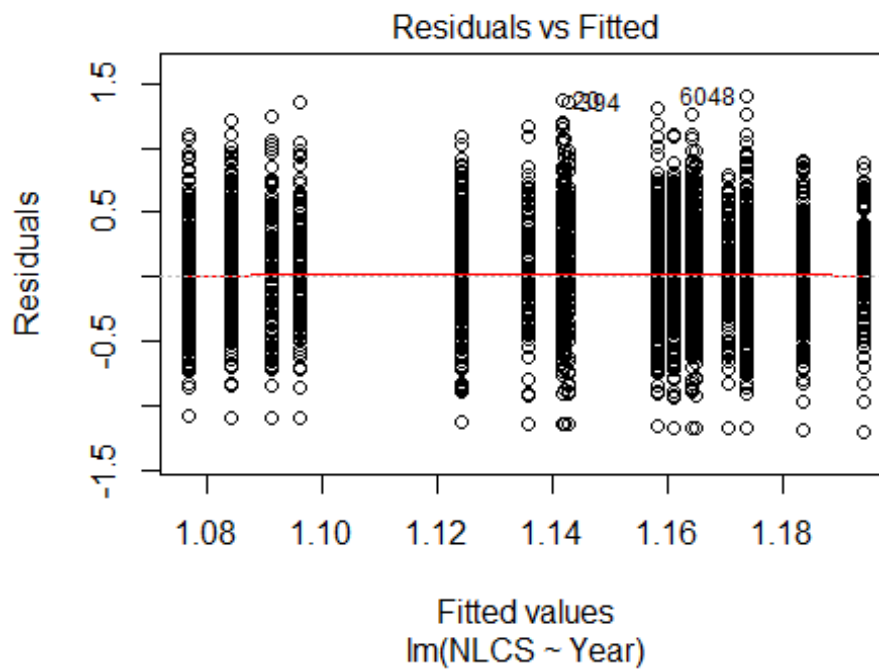
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2474 -0.3278 0.0315 0.3212 1.1822
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1105 0.0929 11.96 <2e-16 ***
## LastAuthorFemale1 -0.0103 0.0453 -0.23 0.8210
## Year1997 -0.0117 0.1580 -0.07 0.9408
## Year1998 -0.1514 0.1536 -0.99 0.3248
## Year1999 0.3375 0.1205 2.80 0.0052 **
## Year2000 0.1195 0.1378 0.87 0.3865
## Year2001 0.0999 0.1198 0.83 0.4047
## Year2002 0.0154 0.1256 0.12 0.9025
## Year2003 0.0540 0.1074 0.50 0.6152
## Year2004 0.0870 0.1124 0.77 0.4392
## Year2005 -0.0831 0.1166 -0.71 0.4763
## Year2006 0.0556 0.1165 0.48 0.6335
```

```

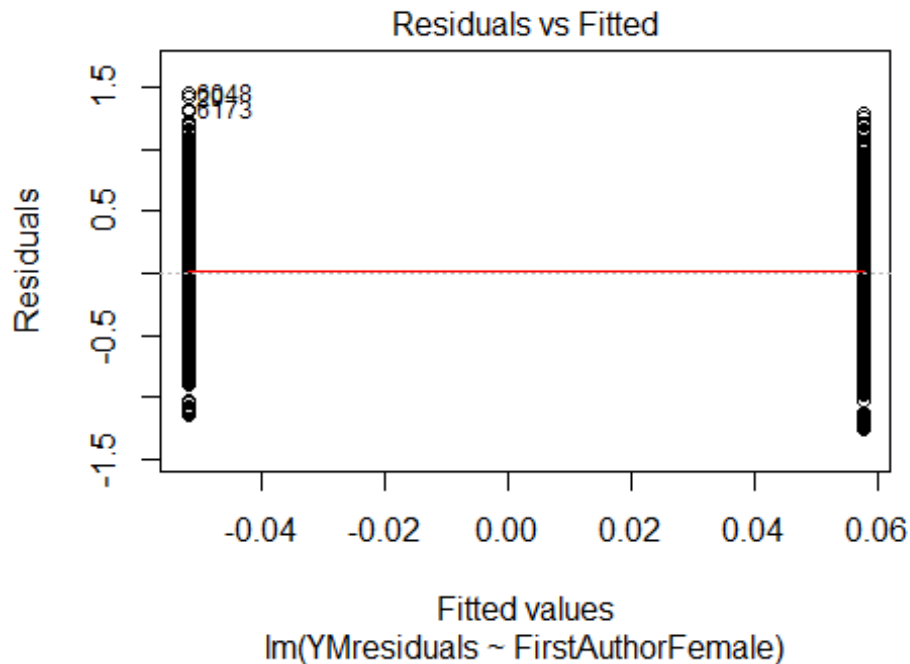
## Year2007          0.1306      0.1193      1.09      0.2740
## Year2008          0.0707      0.1163      0.61      0.5434
## Year2009          0.1063      0.1107      0.96      0.3374
## Year2010          0.1370      0.1199      1.14      0.2536
## Year2011         -0.0367      0.1170     -0.31      0.7540
## Year2012         -0.0118      0.1130     -0.10      0.9168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.476
## Multiple R-squared:  0.0344, Adjusted R-squared:  0.0114
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 667 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.473  0.869  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      1.37e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 732"
## [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
##  246  224  232  217  227  233  212  238  301  303  374  337  373  440  471
## 2011 2012
##  505  472
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  131  131  141  148  117  112  138  145  212  207  271  238  266  305  321
## 2011 2012

```

```
## 376 352
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 118 124 134 105 95 124 124 185 170 243 214 233 268 293
## 2011 2012
## 347 314
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 49, df = 16, p-value = 3e-05
```

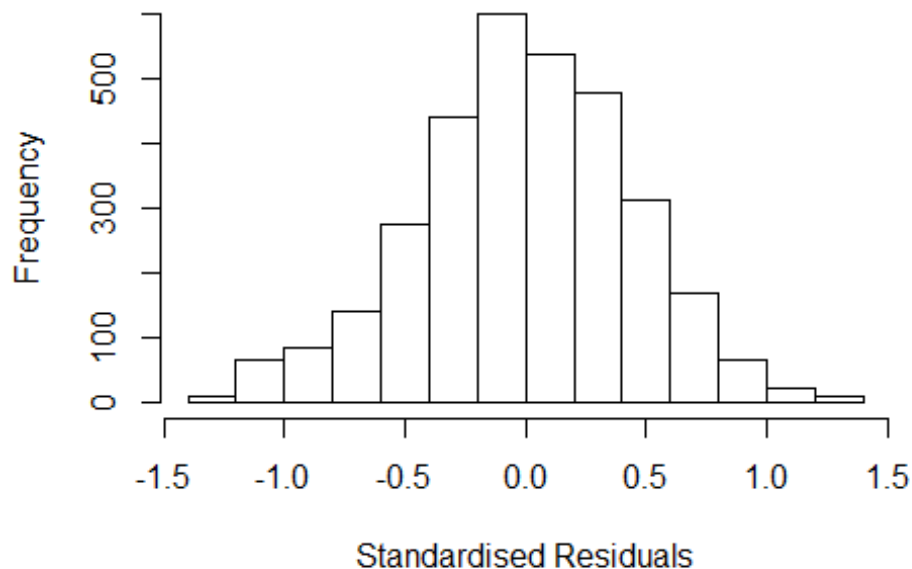


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 31, df = 1, p-value = 2e-08
```



```
## [1] "Female first author team size 2018 geometric mean: 5.44458824394139"
## [1] "Male first author team size 2018 geometric mean: 4.4579845287223"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18000, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.99576793919012"
## [1] "Male last author team size 2018 geometric mean: 4.85464446108478"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, 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 1.112 1      1.055
## LastAuthorFemale  1.101 1      1.049
## UniqueAuthors    1.220 4      1.025
## Year             1.247 16      1.007
```

Residuals from first and last author and team size



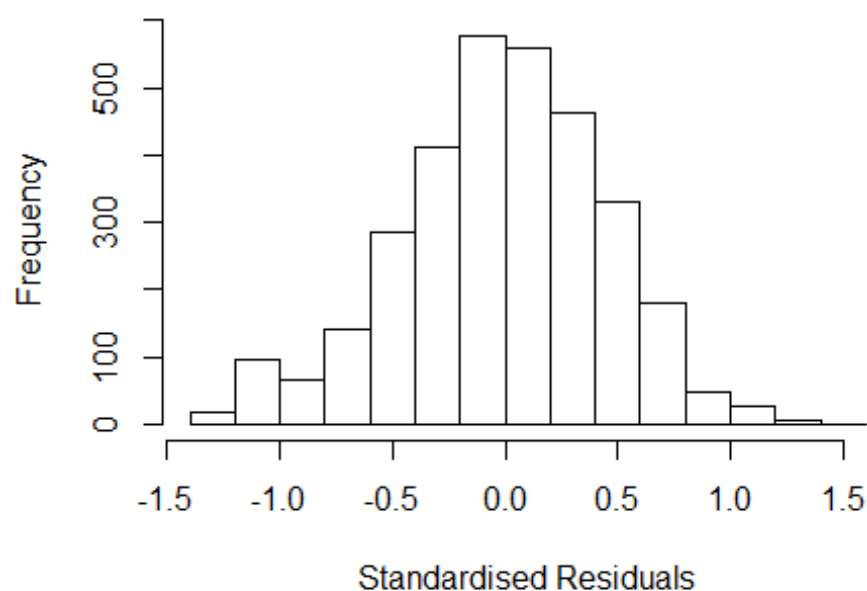
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35266 -0.28476 -0.00366 0.29186 1.38046
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.85517 0.05875 14.56 < 2e-16 ***
## FirstAuthorFemale1 0.06470 0.01670 3.87 0.00011 ***
## LastAuthorFemale1 0.06118 0.01769 3.46 0.00055 ***
## UniqueAuthors2 0.21219 0.04453 4.76 2.0e-06 ***
## UniqueAuthors3 0.26979 0.04181 6.45 1.3e-10 ***
## UniqueAuthors4 0.31103 0.04273 7.28 4.2e-13 ***
## UniqueAuthors5 0.41455 0.04009 10.34 < 2e-16 ***
## Year1997 0.01824 0.07031 0.26 0.79529
## Year1998 -0.03217 0.06933 -0.46 0.64269
## Year1999 -0.07334 0.06335 -1.16 0.24707
```

```

## Year2000          0.00548    0.06177    0.09  0.92935
## Year2001          -0.04530    0.06878   -0.66  0.51018
## Year2002          -0.03550    0.05884   -0.60  0.54631
## Year2003          -0.08312    0.05905   -1.41  0.15936
## Year2004          -0.12952    0.05793   -2.24  0.02545 *
## Year2005          -0.00647    0.05752   -0.11  0.91039
## Year2006          -0.05051    0.05546   -0.91  0.36256
## Year2007          -0.03888    0.05443   -0.71  0.47504
## Year2008          -0.10577    0.05572   -1.90  0.05774 .
## Year2009          -0.07436    0.05536   -1.34  0.17927
## Year2010          -0.04754    0.05388   -0.88  0.37771
## Year2011          -0.03331    0.05265   -0.63  0.52701
## Year2012          -0.06632    0.05427   -1.22  0.22178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0815, Adjusted R-squared:  0.0751
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 258 weights are ~= 1. The remaining 2952 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.286  0.869  0.951  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      3.12e-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 1.096 1      1.047
## LastAuthorFemale  1.074 1      1.037
## Year              1.063 16      1.002

```

Residuals from first and last author



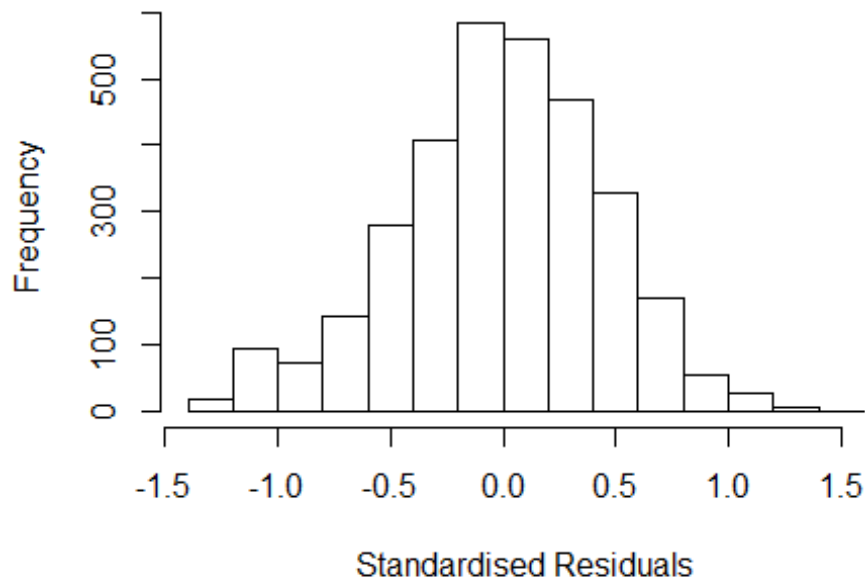
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26371 -0.28965  0.00162  0.29764  1.44683
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11960    0.04941   22.66 < 2e-16 ***
## FirstAuthorFemale1 0.08688    0.01704    5.10 3.6e-07 ***
## LastAuthorFemale1 0.05267    0.01776    2.97 0.003 **
## Year1997         0.02248    0.07541    0.30 0.766
## Year1998        -0.03211    0.06934   -0.46 0.643
## Year1999        -0.08166    0.06515   -1.25 0.210
## Year2000         0.01821    0.06324    0.29 0.773
## Year2001        -0.01355    0.07462   -0.18 0.856
## Year2002        -0.00672    0.05985   -0.11 0.911
## Year2003        -0.04685    0.06047   -0.77 0.439
## Year2004        -0.10406    0.05994   -1.74 0.083 .
## Year2005         0.01278    0.05837    0.22 0.827
```

```

## Year2006      -0.00773    0.05644   -0.14    0.891
## Year2007      -0.00761    0.05587   -0.14    0.892
## Year2008      -0.06954    0.05842   -1.19    0.234
## Year2009      -0.03988    0.05696   -0.70    0.484
## Year2010       0.00209    0.05585    0.04    0.970
## Year2011       0.00457    0.05432    0.08    0.933
## Year2012      -0.01520    0.05594   -0.27    0.786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0198, Adjusted R-squared:  0.0143
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 283 weights are ~= 1. The remaining 2927 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.869  0.949  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      3.12e-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 1.042 1      1.021
## Year              1.042 16      1.001

```


Residuals from first author



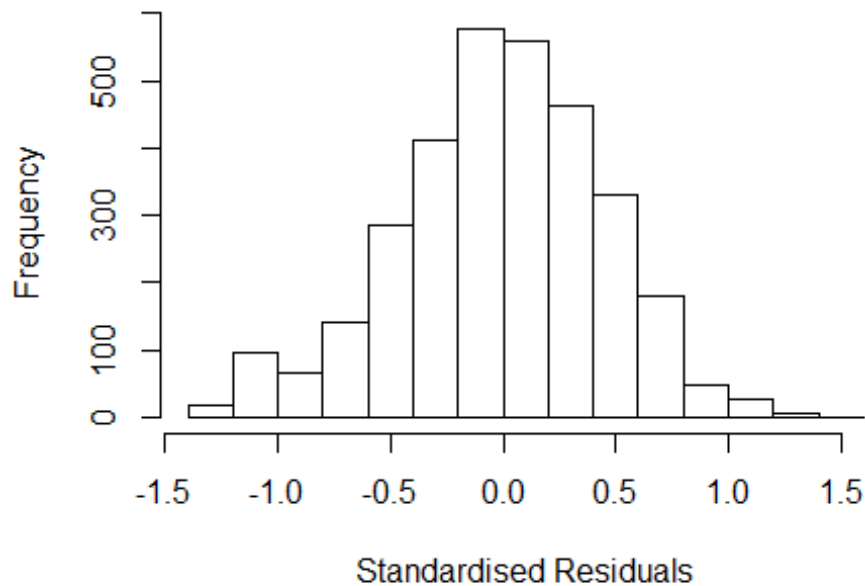
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24767 -0.28889 0.00121 0.30239 1.43339
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12944 0.04934 22.89 < 2e-16 ***
## FirstAuthorFemale1 0.09738 0.01665 5.85 5.4e-09 ***
## Year1997 0.02085 0.07522 0.28 0.782
## Year1998 -0.03343 0.06935 -0.48 0.630
## Year1999 -0.08169 0.06525 -1.25 0.211
## Year2000 0.01984 0.06309 0.31 0.753
## Year2001 -0.01539 0.07474 -0.21 0.837
## Year2002 -0.00976 0.05998 -0.16 0.871
## Year2003 -0.04442 0.06040 -0.74 0.462
## Year2004 -0.10565 0.06003 -1.76 0.079 .
## Year2005 0.01595 0.05846 0.27 0.785
## Year2006 -0.00638 0.05654 -0.11 0.910
```

```

## Year2007          -0.00502    0.05594   -0.09    0.929
## Year2008          -0.06918    0.05855   -1.18    0.237
## Year2009          -0.03745    0.05690   -0.66    0.511
## Year2010           0.00376    0.05592    0.07    0.946
## Year2011           0.00817    0.05437    0.15    0.881
## Year2012          -0.01189    0.05595   -0.21    0.832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0171, Adjusted R-squared:  0.0119
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 289 weights are ~= 1. The remaining 2921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.267  0.868  0.949  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      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 1.022 1          1.011
## Year              1.022 16          1.001

```

Residuals from last author



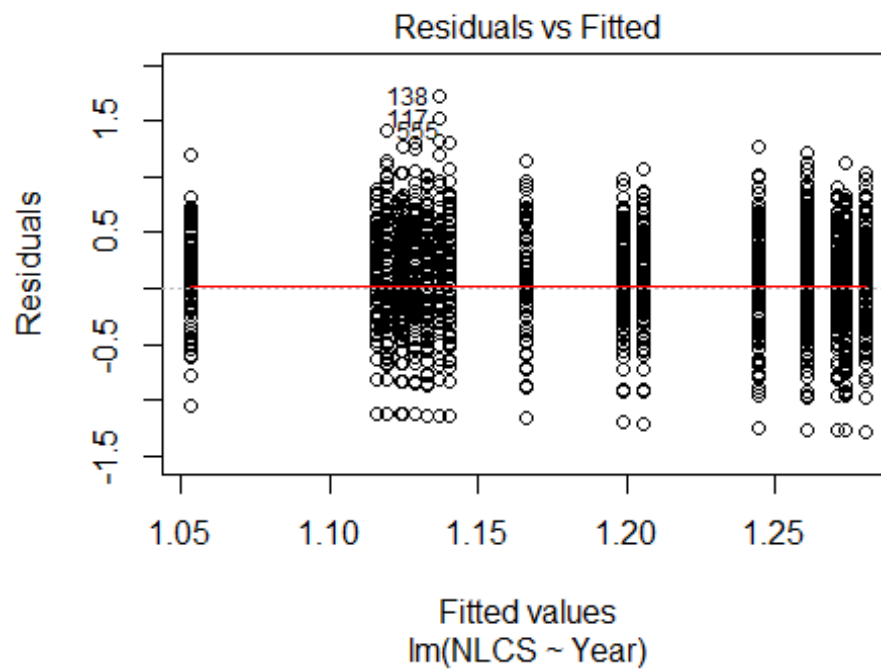
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.25105 -0.28821  0.00508  0.30432  1.41111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14359    0.04921   23.24  < 2e-16 ***
## LastAuthorFemale1 0.07239    0.01728    4.19  2.9e-05 ***
## Year1997         0.03507    0.07563    0.46    0.64
## Year1998        -0.03056    0.06979   -0.44    0.66
## Year1999        -0.07460    0.06561   -1.14    0.26
## Year2000         0.02405    0.06368    0.38    0.71
## Year2001        -0.00273    0.07509   -0.04    0.97
## Year2002         0.00680    0.06017    0.11    0.91
## Year2003        -0.03777    0.06109   -0.62    0.54
## Year2004        -0.08619    0.05991   -1.44    0.15
## Year2005         0.02656    0.05853    0.45    0.65
## Year2006         0.00805    0.05647    0.14    0.89
```

```

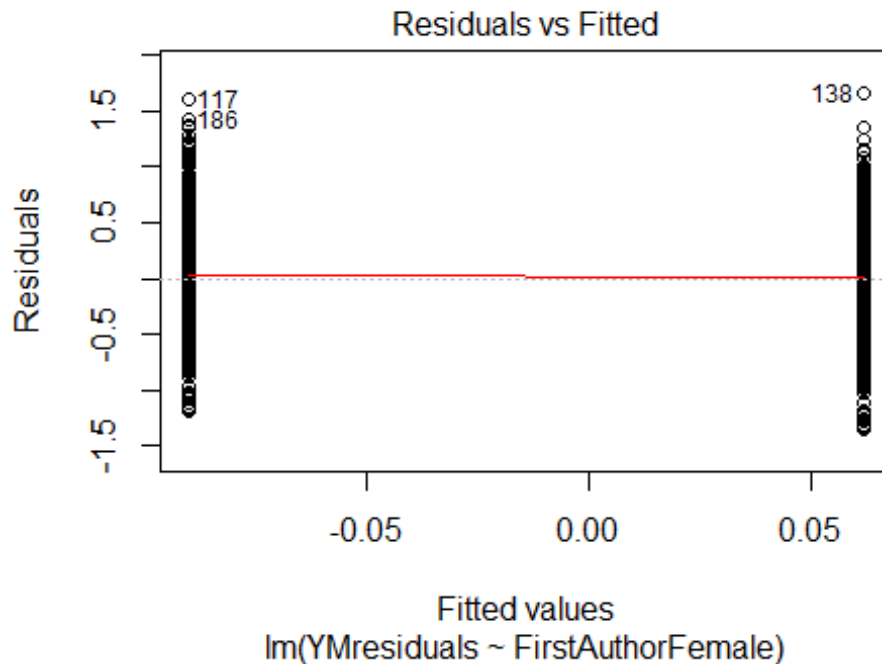
## Year2007      0.00667    0.05601    0.12    0.91
## Year2008     -0.05609    0.05847   -0.96    0.34
## Year2009     -0.02505    0.05701   -0.44    0.66
## Year2010      0.02039    0.05576    0.37    0.71
## Year2011      0.01630    0.05436    0.30    0.76
## Year2012      0.00419    0.05586    0.08    0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00611
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 301 weights are ~= 1. The remaining 2909 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.279  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      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: 3210"
## [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
##  197  194  219  191  214  223  222  189  185  211  222  224  258  268  277
## 2011 2012
##  318  327
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  101   98  105   98  102   92  115  111  134  135  140  154  154  177  171
## 2011 2012

```

```
## 231 226
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 87 95 90 91 82 103 91 119 120 126 141 136 163 154
## 2011 2012
## 203 206
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 50, df = 16, p-value = 2e-05
```

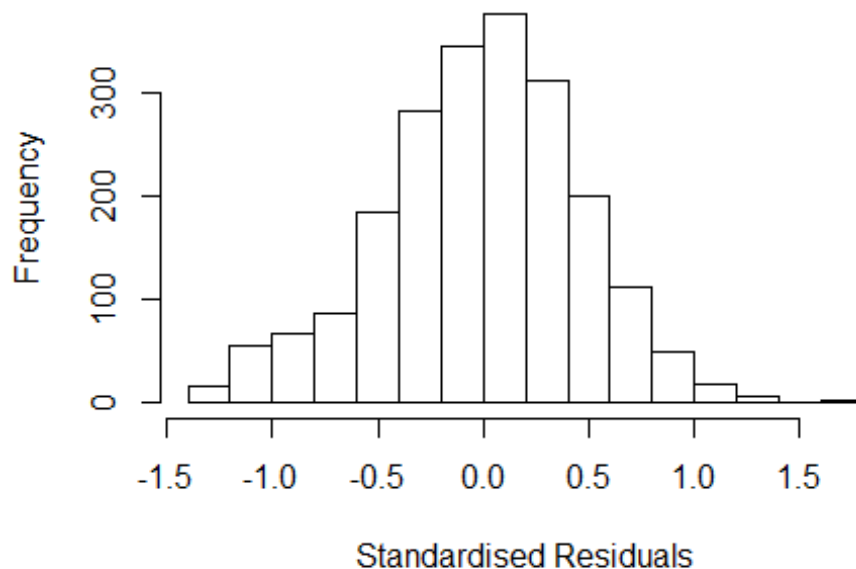


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 4.19800995070241"
## [1] "Male first author team size 2018 geometric mean: 3.68245568565805"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7600, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.91489643499509"
## [1] "Male last author team size 2018 geometric mean: 4.46626216892902"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7400, p-value = 0.004
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.279 1      1.131
## LastAuthorFemale  1.255 1      1.120
## UniqueAuthors    1.223 4      1.025
## Year              1.469 16     1.012
```

Residuals from first and last author and team size



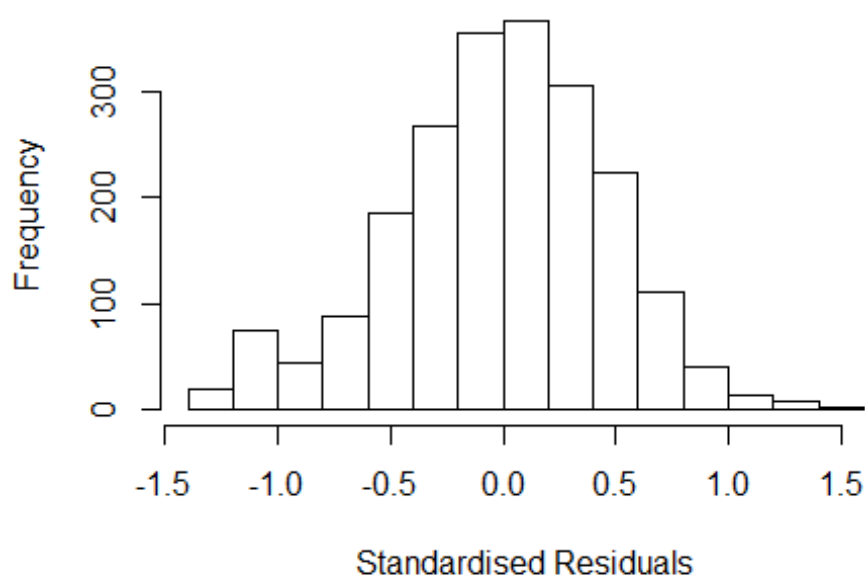
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38585 -0.29945  0.00929  0.30709  1.77249
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8775    0.0702   12.49 < 2e-16 ***
## FirstAuthorFemale1 0.1205    0.0242    4.98 6.7e-07 ***
## LastAuthorFemale1 0.0956    0.0231    4.14 3.6e-05 ***
## UniqueAuthors2    0.2160    0.0506    4.27 2.1e-05 ***
## UniqueAuthors3    0.2133    0.0464    4.60 4.5e-06 ***
## UniqueAuthors4    0.2314    0.0482    4.80 1.7e-06 ***
## UniqueAuthors5    0.3172    0.0453    7.00 3.4e-12 ***
## Year1997          0.0253    0.0814    0.31  0.76
## Year1998         -0.0332    0.0860   -0.39  0.70
## Year1999         -0.0367    0.0811   -0.45  0.65
```

```

## Year2000          -0.0213      0.0874   -0.24      0.81
## Year2001          -0.0531      0.0823   -0.65      0.52
## Year2002          -0.1285      0.0793   -1.62      0.11
## Year2003          -0.0704      0.0730   -0.96      0.34
## Year2004          -0.0645      0.0725   -0.89      0.37
## Year2005          -0.0758      0.0798   -0.95      0.34
## Year2006           0.0111      0.0744     0.15      0.88
## Year2007           0.0763      0.0735     1.04      0.30
## Year2008           0.0162      0.0741     0.22      0.83
## Year2009          -0.0287      0.0696   -0.41      0.68
## Year2010           0.0340      0.0717     0.47      0.64
## Year2011           0.0145      0.0691     0.21      0.83
## Year2012           0.0239      0.0699     0.34      0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.452
## Multiple R-squared:  0.0883, Adjusted R-squared:  0.0786
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 1924 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0889 0.8680 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      4.76e-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 1.233 1      1.111
## LastAuthorFemale  1.224 1      1.106
## Year              1.210 16      1.006

```


Residuals from first and last author



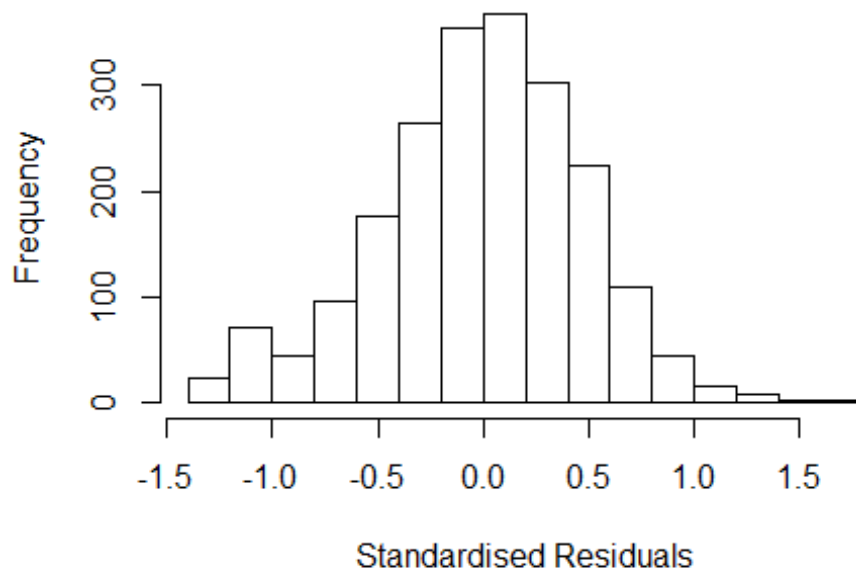
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3943 -0.3043  0.0123  0.3159  1.5819
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0681    0.0586   18.24 < 2e-16 ***
## FirstAuthorFemale1  0.1394    0.0245    5.69 1.4e-08 ***
## LastAuthorFemale1  0.0821    0.0231    3.55 0.00039 ***
## Year1997          0.0242    0.0804    0.30 0.76366
## Year1998         -0.0146    0.0851   -0.17 0.86427
## Year1999         -0.0281    0.0812   -0.35 0.72979
## Year2000         -0.0133    0.0872   -0.15 0.87888
## Year2001         -0.0080    0.0823   -0.10 0.92258
## Year2002         -0.0894    0.0785   -1.14 0.25478
## Year2003         -0.0341    0.0720   -0.47 0.63579
## Year2004         -0.0440    0.0713   -0.62 0.53772
## Year2005         -0.0421    0.0789   -0.53 0.59307
```

```

## Year2006          0.0352      0.0719      0.49  0.62496
## Year2007          0.1048      0.0719      1.46  0.14506
## Year2008          0.0453      0.0722      0.63  0.53044
## Year2009          0.0158      0.0679      0.23  0.81571
## Year2010          0.0798      0.0698      1.14  0.25306
## Year2011          0.0620      0.0674      0.92  0.35771
## Year2012          0.0814      0.0679      1.20  0.23051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0553, Adjusted R-squared:  0.0471
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 181 weights are ~= 1. The remaining 1921 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.201  0.869  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      4.76e-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 1.123 1          1.060
## Year              1.123 16          1.004

```

Residuals from first author



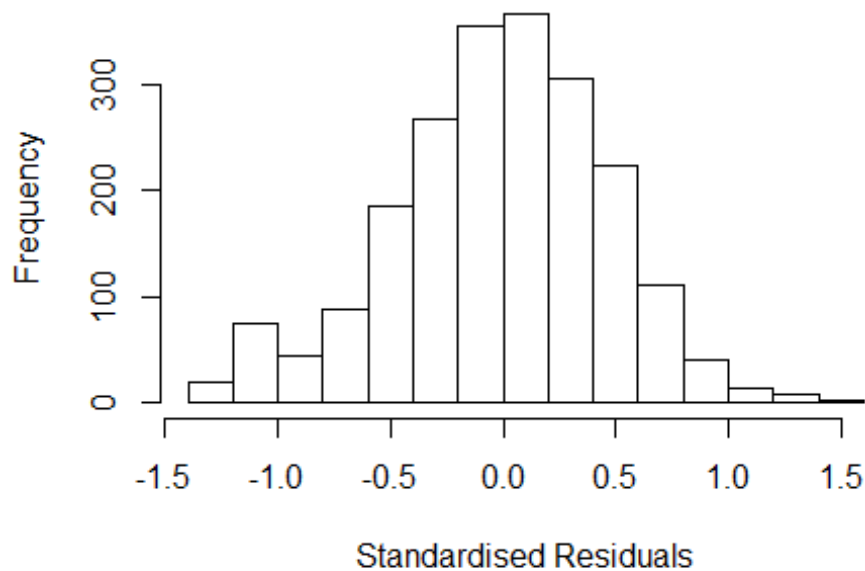
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3600 -0.3105 0.0104 0.3147 1.6065
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0829 0.0595 18.20 <2e-16 ***
## FirstAuthorFemale1 0.1646 0.0235 7.02 3e-12 ***
## Year1997 0.0262 0.0815 0.32 0.75
## Year1998 -0.0217 0.0858 -0.25 0.80
## Year1999 -0.0373 0.0817 -0.46 0.65
## Year2000 -0.0178 0.0880 -0.20 0.84
## Year2001 -0.0205 0.0835 -0.25 0.81
## Year2002 -0.0951 0.0794 -1.20 0.23
## Year2003 -0.0380 0.0729 -0.52 0.60
## Year2004 -0.0452 0.0725 -0.62 0.53
## Year2005 -0.0405 0.0796 -0.51 0.61
## Year2006 0.0416 0.0728 0.57 0.57
```

```

## Year2007          0.1125      0.0727      1.55      0.12
## Year2008          0.0588      0.0728      0.81      0.42
## Year2009          0.0237      0.0688      0.34      0.73
## Year2010          0.0907      0.0706      1.28      0.20
## Year2011          0.0765      0.0680      1.13      0.26
## Year2012          0.0953      0.0687      1.39      0.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.0493, Adjusted R-squared:  0.0415
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 186 weights are ~= 1. The remaining 1916 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  0.869  0.948  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      4.76e-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 1.116 1      1.057
## Year              1.116 16      1.003

```

Residuals from last author



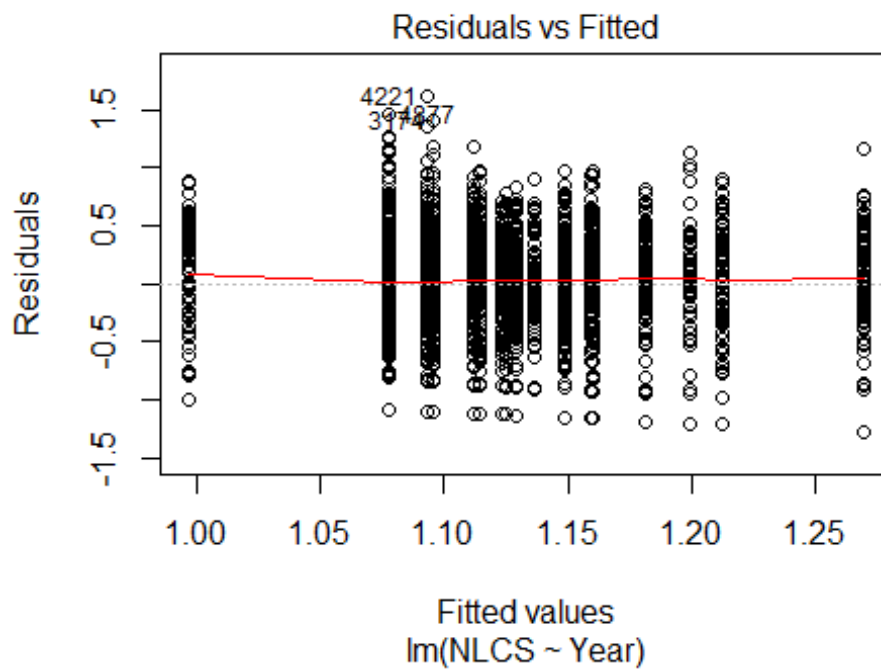
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37604 -0.30552  0.00953  0.30901  1.63082
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10004    0.05844   18.82  < 2e-16 ***
## LastAuthorFemale1 0.12314    0.02216    5.56  3.1e-08 ***
## Year1997        0.02739    0.08069    0.34   0.734
## Year1998        0.00915    0.08687    0.11   0.916
## Year1999        0.00522    0.08097    0.06   0.949
## Year2000        0.02390    0.08915    0.27   0.789
## Year2001        0.02647    0.08279    0.32   0.749
## Year2002       -0.07005    0.07862   -0.89   0.373
## Year2003        0.00885    0.07175    0.12   0.902
## Year2004       -0.01271    0.07103   -0.18   0.858
## Year2005       -0.00884    0.07845   -0.11   0.910
## Year2006        0.07125    0.07158    1.00   0.320
```

```

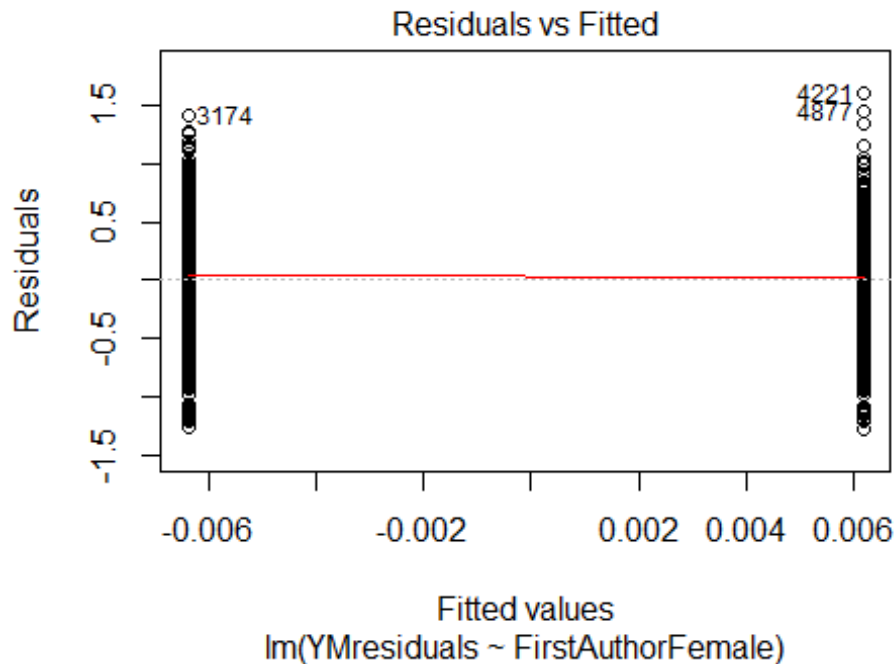
## Year2007      0.15286    0.07096    2.15    0.031 *
## Year2008      0.09274    0.07109    1.30    0.192
## Year2009      0.06887    0.06712    1.03    0.305
## Year2010      0.12231    0.06940    1.76    0.078 .
## Year2011      0.10281    0.06709    1.53    0.126
## Year2012      0.12832    0.06664    1.93    0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.0387, Adjusted R-squared:  0.0309
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 192 weights are ~= 1. The remaining 1910 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.175  0.867  0.949  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      4.76e-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: 2102"
## [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
## 140 144 162 139 172 210 171 152 196 209 261 214 302 281 328
## 2011 2012
## 386 439
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 70 90 90 77 101 100 102 139 150 180 146 189 174 224
## 2011 2012

```

```
## 276 317
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 62 77 79 74 89 87 81 125 130 161 132 164 153 201
## 2011 2012
## 242 288
## [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 = 7e-07
```

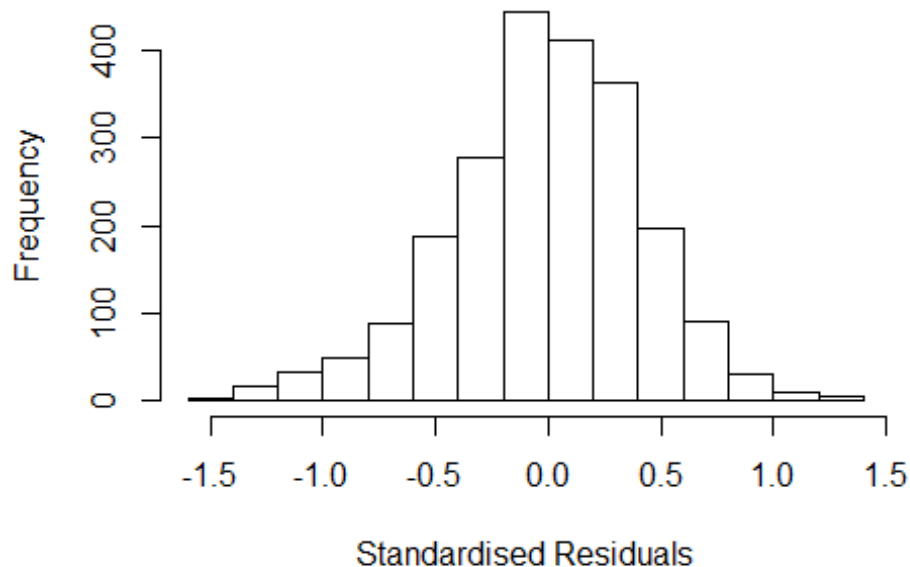


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 17, df = 1, p-value = 4e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 6.375625364017"
## [1] "Male first author team size 2018 geometric mean: 4.79934684604943"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 40000, p-value = 2e-12
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.17290632050246"
## [1] "Male last author team size 2018 geometric mean: 5.20337849439826"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 28000, p-value = 3e-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 1.081 1      1.040
## LastAuthorFemale  1.064 1      1.032
## UniqueAuthors    1.144 4      1.017
## Year              1.235 16     1.007
```


Residuals from first and last author and team size



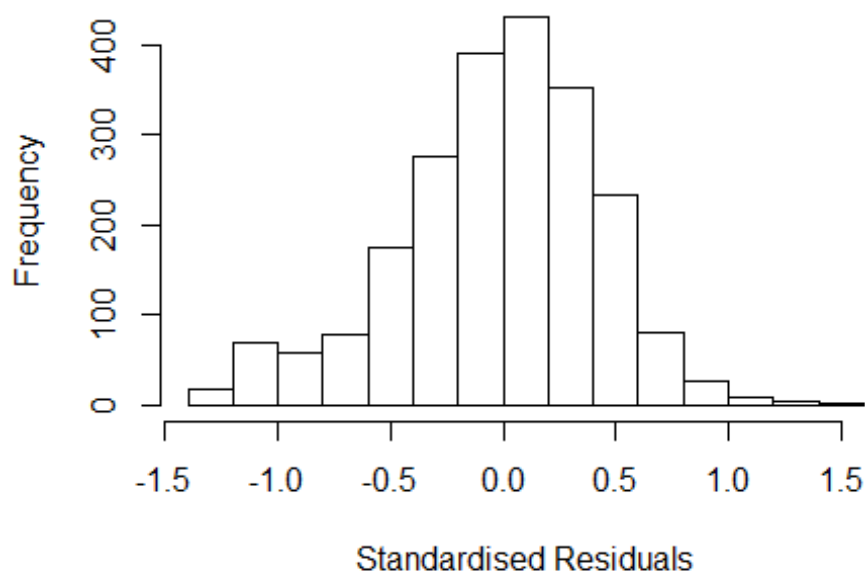
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46372 -0.26923 0.00227 0.26833 1.38370
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.795279 0.076795 10.36 < 2e-16 ***
## FirstAuthorFemale1 0.000967 0.018631 0.05 0.9586
## LastAuthorFemale1 0.011288 0.020032 0.56 0.5732
## UniqueAuthors2 0.386303 0.068228 5.66 1.7e-08 ***
## UniqueAuthors3 0.517295 0.061758 8.38 < 2e-16 ***
## UniqueAuthors4 0.505084 0.060315 8.37 < 2e-16 ***
## UniqueAuthors5 0.611310 0.058016 10.54 < 2e-16 ***
## Year1997 -0.086731 0.084050 -1.03 0.3022
## Year1998 -0.098277 0.081232 -1.21 0.2265
## Year1999 0.045846 0.071928 0.64 0.5239
```

```

## Year2000      -0.208019    0.083024    -2.51    0.0123 *
## Year2001      -0.107057    0.080090    -1.34    0.1815
## Year2002      -0.156407    0.078087    -2.00    0.0453 *
## Year2003      -0.180378    0.070350    -2.56    0.0104 *
## Year2004      -0.153578    0.068325    -2.25    0.0247 *
## Year2005      -0.202052    0.067536    -2.99    0.0028 **
## Year2006      -0.210347    0.066835    -3.15    0.0017 **
## Year2007      -0.191621    0.069203    -2.77    0.0057 **
## Year2008      -0.197762    0.070869    -2.79    0.0053 **
## Year2009      -0.219972    0.070722    -3.11    0.0019 **
## Year2010      -0.195783    0.066832    -2.93    0.0034 **
## Year2011      -0.205501    0.066152    -3.11    0.0019 **
## Year2012      -0.257253    0.065920    -3.90    9.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.123, Adjusted R-squared:  0.114
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 166 weights are ~= 1. The remaining 2036 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.149  0.869  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      4.54e-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 1.078 1      1.038
## LastAuthorFemale  1.053 1      1.026
## Year              1.097 16      1.003

```

Residuals from first and last author



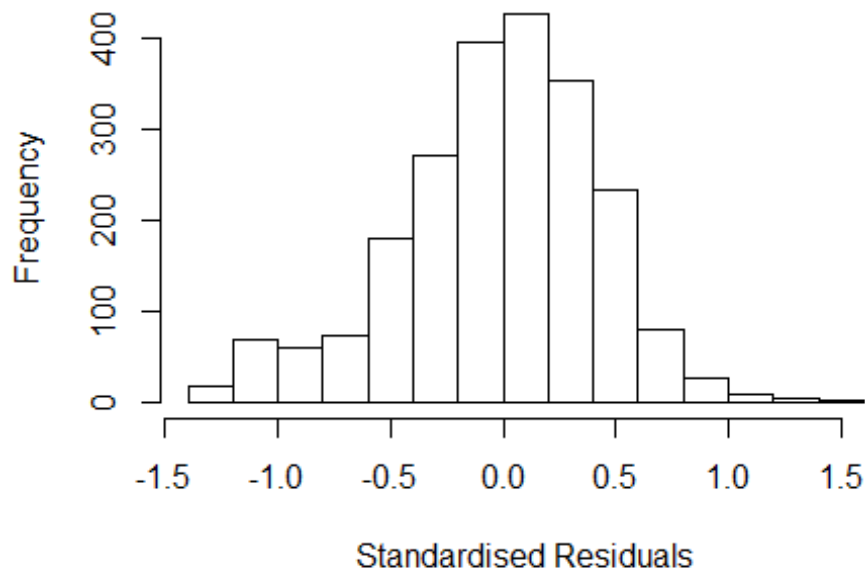
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3342 -0.2806  0.0169  0.2863  1.4298
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23809    0.07027   17.62  <2e-16 ***
## FirstAuthorFemale1  0.01417    0.01953    0.73   0.468
## LastAuthorFemale1 -0.01175    0.02052   -0.57   0.567
## Year1997         0.00472    0.09391    0.05   0.960
## Year1998        -0.03284    0.08752   -0.38   0.708
## Year1999         0.08195    0.08150    1.01   0.315
## Year2000        -0.17908    0.10740   -1.67   0.096 .
## Year2001        -0.02837    0.08656   -0.33   0.743
## Year2002        -0.07107    0.08551   -0.83   0.406
## Year2003        -0.10069    0.07849   -1.28   0.200
## Year2004        -0.06091    0.07776   -0.78   0.434
## Year2005        -0.09714    0.07645   -1.27   0.204
```

```

## Year2006          -0.11405      0.07636      -1.49      0.135
## Year2007          -0.08385      0.07835      -1.07      0.285
## Year2008          -0.08331      0.08031      -1.04      0.300
## Year2009          -0.12563      0.07959      -1.58      0.115
## Year2010          -0.07719      0.07603      -1.02      0.310
## Year2011          -0.10837      0.07641      -1.42      0.156
## Year2012          -0.14806      0.07574      -1.95      0.051 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.00706
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 2019 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.190  0.870  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      4.54e-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 1.065 1          1.032
## Year              1.065 16          1.002

```

Residuals from first author



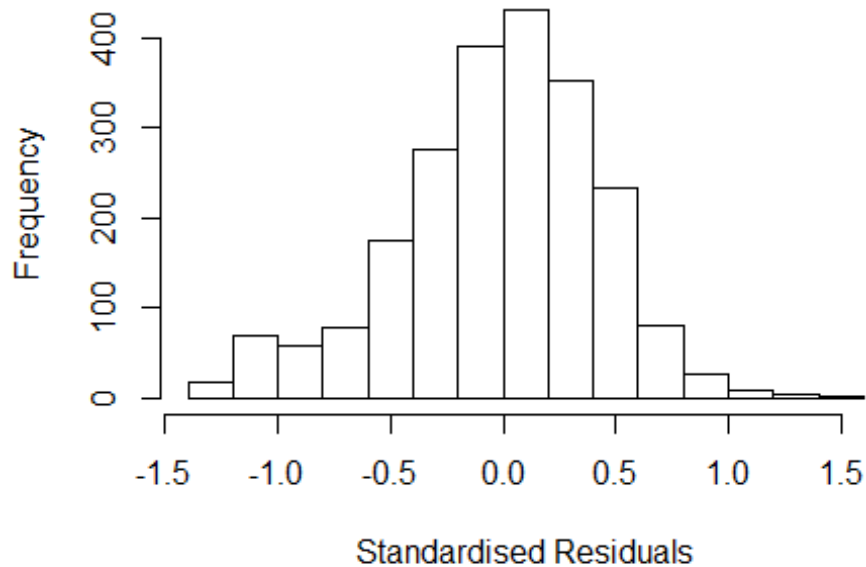
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3299 -0.2812  0.0171  0.2862  1.4342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2359    0.0701   17.64  <2e-16 ***
## FirstAuthorFemale1  0.0124    0.0194    0.64   0.525
## Year1997          0.0054    0.0939    0.06   0.954
## Year1998         -0.0325    0.0873   -0.37   0.710
## Year1999          0.0816    0.0814    1.00   0.316
## Year2000         -0.1793    0.1074   -1.67   0.095 .
## Year2001         -0.0288    0.0865   -0.33   0.739
## Year2002         -0.0715    0.0854   -0.84   0.403
## Year2003         -0.1012    0.0785   -1.29   0.197
## Year2004         -0.0607    0.0777   -0.78   0.435
## Year2005         -0.0972    0.0764   -1.27   0.203
## Year2006         -0.1142    0.0763   -1.50   0.135
```

```

## Year2007          -0.0847      0.0782   -1.08    0.279
## Year2008          -0.0845      0.0801   -1.05    0.292
## Year2009          -0.1258      0.0795   -1.58    0.114
## Year2010          -0.0780      0.0759   -1.03    0.304
## Year2011          -0.1096      0.0762   -1.44    0.151
## Year2012          -0.1485      0.0757   -1.96    0.050 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00737
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 2018 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.186  0.870  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      4.54e-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 1.04 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
## -1.3262 -0.2823 0.0151 0.2848 1.4368
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24182 0.07031 17.66 <2e-16 ***
## LastAuthorFemale1 -0.00924 0.02041 -0.45 0.651
## Year1997 0.00772 0.09379 0.08 0.934
## Year1998 -0.03254 0.08765 -0.37 0.710
## Year1999 0.08442 0.08148 1.04 0.300
## Year2000 -0.17638 0.10705 -1.65 0.100 .
## Year2001 -0.02503 0.08634 -0.29 0.772
## Year2002 -0.06874 0.08563 -0.80 0.422
## Year2003 -0.09725 0.07841 -1.24 0.215
## Year2004 -0.05816 0.07774 -0.75 0.454
## Year2005 -0.09387 0.07644 -1.23 0.220
## Year2006 -0.11190 0.07638 -1.46 0.143
```

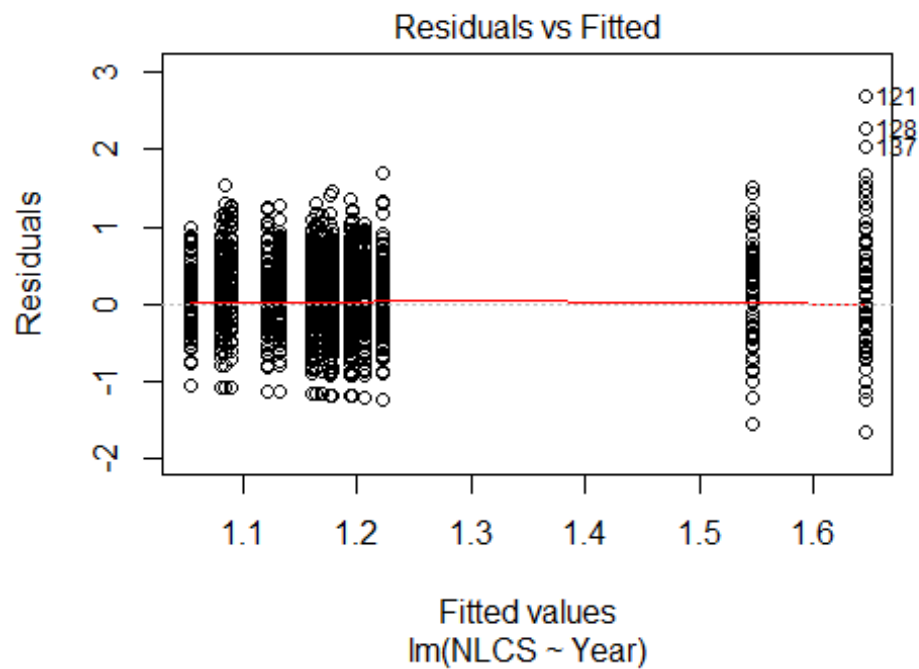
```

## Year2007          -0.07991      0.07821    -1.02      0.307
## Year2008          -0.07917      0.07995    -0.99      0.322
## Year2009          -0.12266      0.07957    -1.54      0.123
## Year2010          -0.07356      0.07583    -0.97      0.332
## Year2011          -0.10531      0.07632    -1.38      0.168
## Year2012          -0.14463      0.07563    -1.91      0.056 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.00721
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 190 weights are ~= 1. The remaining 2012 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.186  0.870  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      4.54e-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: 2202"
## [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
## 182 220 188 173 171 176 150 172 192 193 252 260 202 244 227
## 2011 2012
## 203 214
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 77 81 108 110 108 98 118 127 141 147 179 193 144 165 162
## 2011 2012

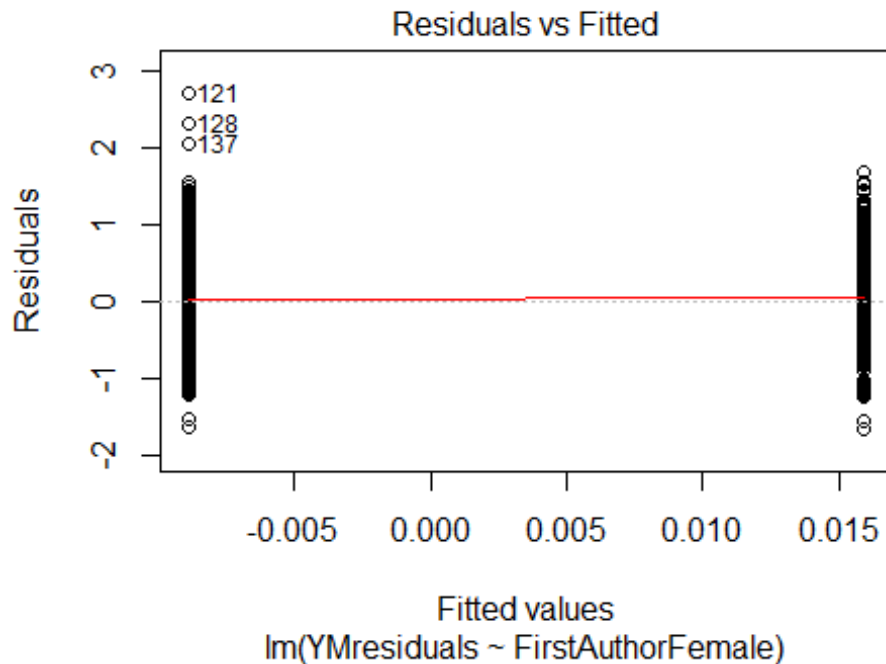
```



```
## 155 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 72 68 87 98 90 83 91 112 125 128 150 158 125 132 124
## 2011 2012
## 130 125
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```

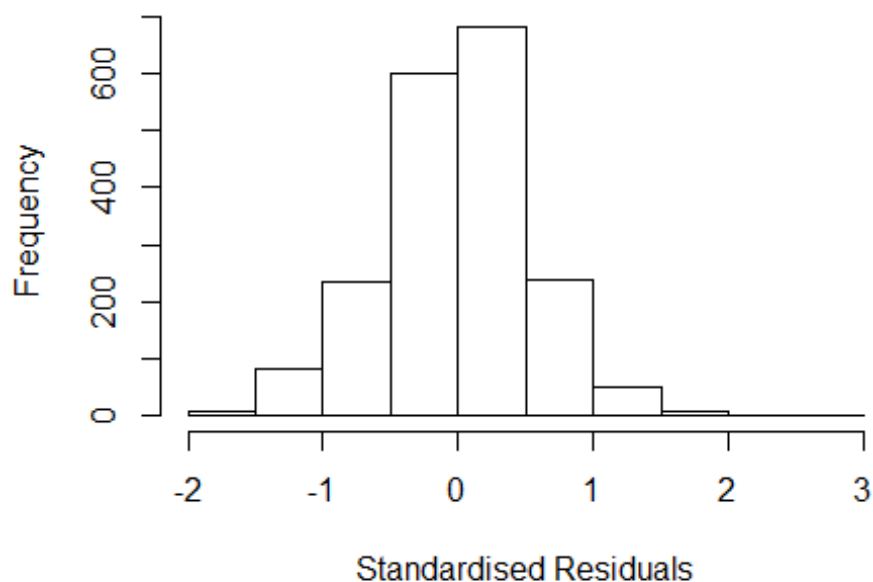


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.46418272147203"
## [1] "Male first author team size 2018 geometric mean: 3.65095047881641"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.08387338144055"
## [1] "Male last author team size 2018 geometric mean: 3.99678902457644"
##
## 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 1.100 1          1.049
## LastAuthorFemale  1.057 1          1.028
## UniqueAuthors    1.338 4          1.037
## Year              1.363 16         1.010
```

Residuals from first and last author and team size



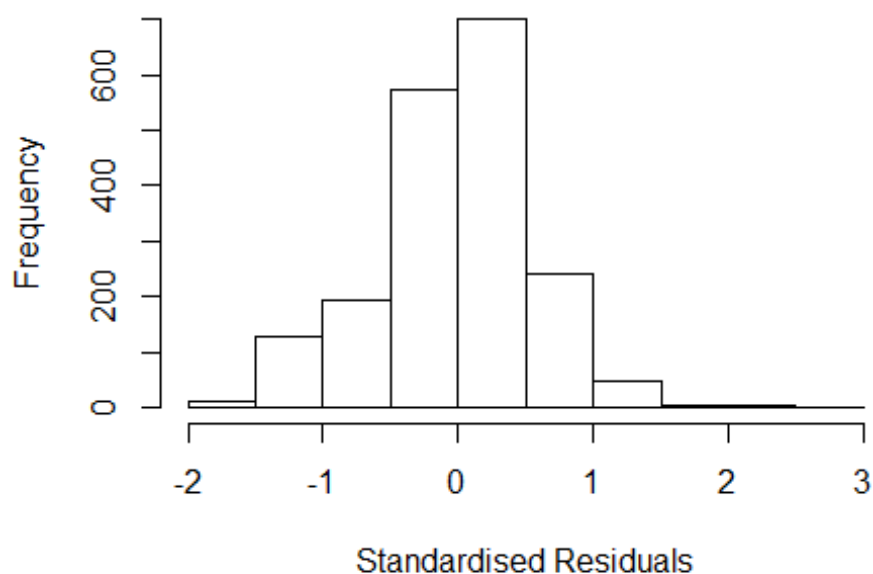
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 121 0029842865 4.334 1996      2731      1      2.722
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5858 -0.3280  0.0192  0.3435  2.7223
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2279    0.1407   8.73  < 2e-16 ***
## FirstAuthorFemale1 -0.0293    0.0254  -1.15  0.24879
## LastAuthorFemale1 -0.0423    0.0298  -1.42  0.15616
## UniqueAuthors2      0.3112    0.0560   5.56  3.1e-08 ***
## UniqueAuthors3      0.3838    0.0550   6.98  4.2e-12 ***
## UniqueAuthors4      0.4771    0.0557   8.56  < 2e-16 ***
## UniqueAuthors5      0.5905    0.0550  10.74  < 2e-16 ***
## Year1997           0.0165    0.1723   0.10  0.92381
## Year1998          -0.4931    0.1456  -3.39  0.00072 ***
## Year1999          -0.4363    0.1419  -3.07  0.00214 **
```

```

## Year2000          -0.3639      0.1537    -2.37  0.01804 *
## Year2001          -0.4783      0.1444    -3.31  0.00095 ***
## Year2002          -0.4155      0.1435    -2.89  0.00384 **
## Year2003          -0.4181      0.1424    -2.94  0.00335 **
## Year2004          -0.4352      0.1390    -3.13  0.00177 **
## Year2005          -0.5221      0.1388    -3.76  0.00017 ***
## Year2006          -0.4512      0.1380    -3.27  0.00109 **
## Year2007          -0.4416      0.1379    -3.20  0.00138 **
## Year2008          -0.3629      0.1409    -2.58  0.01007 *
## Year2009          -0.4889      0.1403    -3.49  0.00050 ***
## Year2010          -0.4568      0.1421    -3.22  0.00132 **
## Year2011          -0.4574      0.1389    -3.29  0.00101 **
## Year2012          -0.4949      0.1410    -3.51  0.00046 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.493
## Multiple R-squared:  0.136, Adjusted R-squared:  0.126
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 2 observations c(26,30) are outliers with |weight| = 0 ( < 5.3e-05);
## 163 weights are ~= 1. The remaining 1733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.102  0.865  0.948  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           5.27e-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 1.045 1 1.022
## LastAuthorFemale 1.047 1 1.023
## Year 1.082 16 1.002

```

Residuals from first and last author



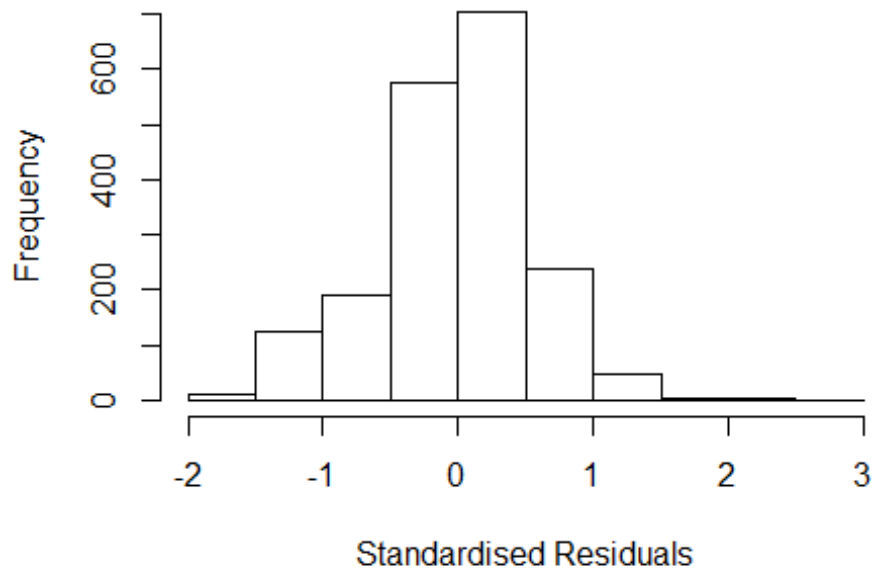
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 121 0029842865 4.334 1996      2731      1      2.775
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6227 -0.3353  0.0245  0.3466  2.7751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5589     0.1509   10.33  <2e-16 ***
## FirstAuthorFemale1  0.0377     0.0259    1.45   0.1464
## LastAuthorFemale1 -0.0444     0.0305   -1.46   0.1454
## Year1997          0.0261     0.1835    0.14   0.8868
## Year1998         -0.4966     0.1631   -3.04   0.0024 **
## Year1999         -0.4791     0.1585   -3.02   0.0025 **
## Year2000         -0.4188     0.1683   -2.49   0.0129 *
## Year2001         -0.4784     0.1641   -2.92   0.0036 **
## Year2002         -0.3821     0.1612   -2.37   0.0179 *
## Year2003         -0.4084     0.1591   -2.57   0.0104 *
## Year2004         -0.3807     0.1558   -2.44   0.0146 *
## Year2005         -0.4781     0.1556   -3.07   0.0021 **
```

```

## Year2006          -0.3786      0.1544    -2.45    0.0143 *
## Year2007          -0.4159      0.1545    -2.69    0.0072 **
## Year2008          -0.3169      0.1568    -2.02    0.0434 *
## Year2009          -0.4333      0.1566    -2.77    0.0057 **
## Year2010          -0.3695      0.1587    -2.33    0.0200 *
## Year2011          -0.3847      0.1548    -2.48    0.0130 *
## Year2012          -0.4062      0.1575    -2.58    0.0100 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.5
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0364
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 2 observations c(26,30) are outliers with |weight| = 0 ( < 5.3e-05);
## 173 weights are ~ = 1. The remaining 1723 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0209 0.8540 0.9480 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          5.27e-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 1.043 1          1.021
## Year              1.043 16          1.001

```

Residuals from first author



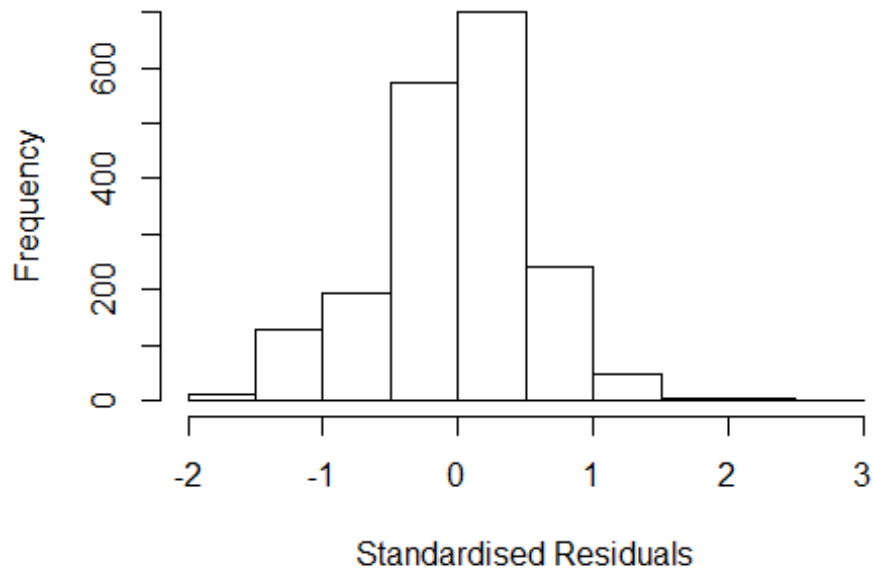
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 121 0029842865 4.334 1996      2731      1      2.775
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6050 -0.3345  0.0244  0.3345  2.7844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5496     0.1519   10.20  <2e-16 ***
## FirstAuthorFemale1  0.0316     0.0260    1.21   0.2252
## Year1997          0.0239     0.1844    0.13   0.8971
## Year1998         -0.4932     0.1646   -3.00   0.0028 **
## Year1999         -0.4799     0.1600   -3.00   0.0027 **
## Year2000         -0.4149     0.1695   -2.45   0.0145 *
## Year2001         -0.4760     0.1656   -2.88   0.0041 **
## Year2002         -0.3762     0.1624   -2.32   0.0207 *
## Year2003         -0.4071     0.1605   -2.54   0.0113 *
## Year2004         -0.3780     0.1571   -2.41   0.0162 *
## Year2005         -0.4775     0.1570   -3.04   0.0024 **
## Year2006         -0.3773     0.1559   -2.42   0.0156 *
```

```

## Year2007          -0.4151      0.1560    -2.66    0.0079 **
## Year2008          -0.3139      0.1581    -1.99    0.0472 *
## Year2009          -0.4332      0.1580    -2.74    0.0062 **
## Year2010          -0.3700      0.1601    -2.31    0.0210 *
## Year2011          -0.3849      0.1564    -2.46    0.0139 *
## Year2012          -0.4071      0.1590    -2.56    0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.5
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0356
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 2 observations c(26,30) are outliers with |weight| = 0 ( < 5.3e-05);
## 178 weights are ~= 1. The remaining 1718 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0292 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      5.27e-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 1.046 1          1.023
## Year          1.046 16          1.001

```


Residuals from last author



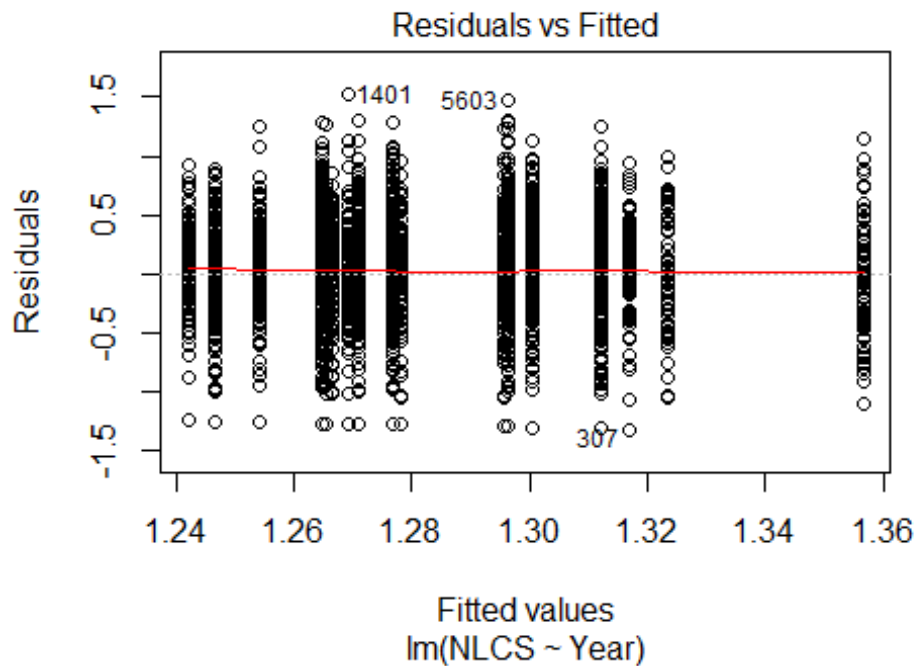
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 121 0029842865 4.334 1996      2731      1      2.775
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5945 -0.3357  0.0293  0.3400  2.7614
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5726     0.1510   10.41  <2e-16 ***
## LastAuthorFemale1 -0.0377     0.0305   -1.23   0.2178
## Year1997          0.0219     0.1841    0.12   0.9055
## Year1998         -0.5001     0.1636   -3.06   0.0023 **
## Year1999         -0.4808     0.1590   -3.02   0.0025 **
## Year2000         -0.4233     0.1688   -2.51   0.0123 *
## Year2001         -0.4810     0.1646   -2.92   0.0035 **
## Year2002         -0.3841     0.1619   -2.37   0.0177 *
## Year2003         -0.4086     0.1597   -2.56   0.0106 *
## Year2004         -0.3840     0.1562   -2.46   0.0141 *
## Year2005         -0.4809     0.1560   -3.08   0.0021 **
## Year2006         -0.3789     0.1549   -2.45   0.0145 *
```

```

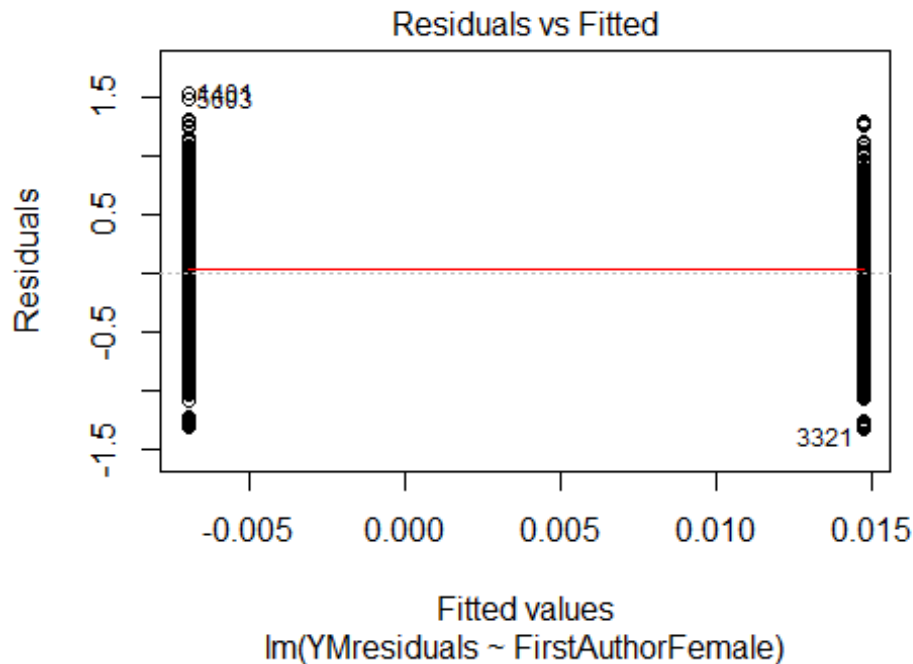
## Year2007          -0.4143      0.1551   -2.67    0.0076 **
## Year2008          -0.3186      0.1573   -2.03    0.0429 *
## Year2009          -0.4317      0.1571   -2.75    0.0061 **
## Year2010          -0.3675      0.1592   -2.31    0.0211 *
## Year2011          -0.3839      0.1552   -2.47    0.0135 *
## Year2012          -0.4082      0.1579   -2.58    0.0098 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.5
## Multiple R-squared:  0.0445, Adjusted R-squared:  0.0359
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 2 observations c(26,30) are outliers with |weight| = 0 ( < 5.3e-05);
## 177 weights are ~= 1. The remaining 1719 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0255 0.8550 0.9480 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      5.27e-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: 1898"
## [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
## 117 121 155 167 167 224 187 187 240 292 296 354 363 365 448
## 2011 2012
## 445 461
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 71 87 100 99 117 118 120 132 187 176 267 267 263 334

```

```
## 2011 2012
## 338 358
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 68 79 96 86 105 108 109 121 162 157 242 234 240 302
## 2011 2012
## 313 326
## [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
```

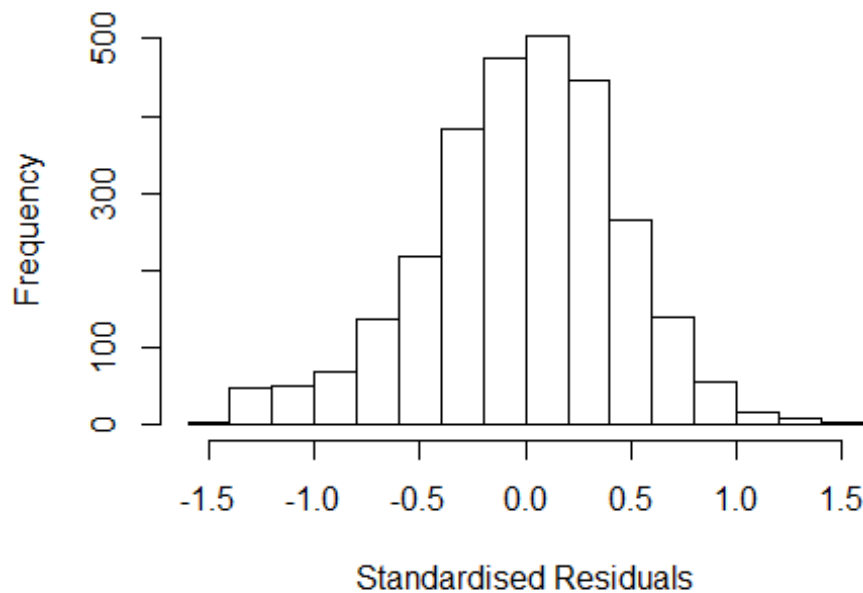


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.11191428782139"
## [1] "Male first author team size 2018 geometric mean: 4.35845098751694"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.90840078529671"
## [1] "Male last author team size 2018 geometric mean: 4.40141332548126"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, 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 1.092 1          1.045
## LastAuthorFemale  1.077 1          1.038
## UniqueAuthors     1.126 4          1.015
## Year              1.164 16         1.005
```

Residuals from first and last author and team size



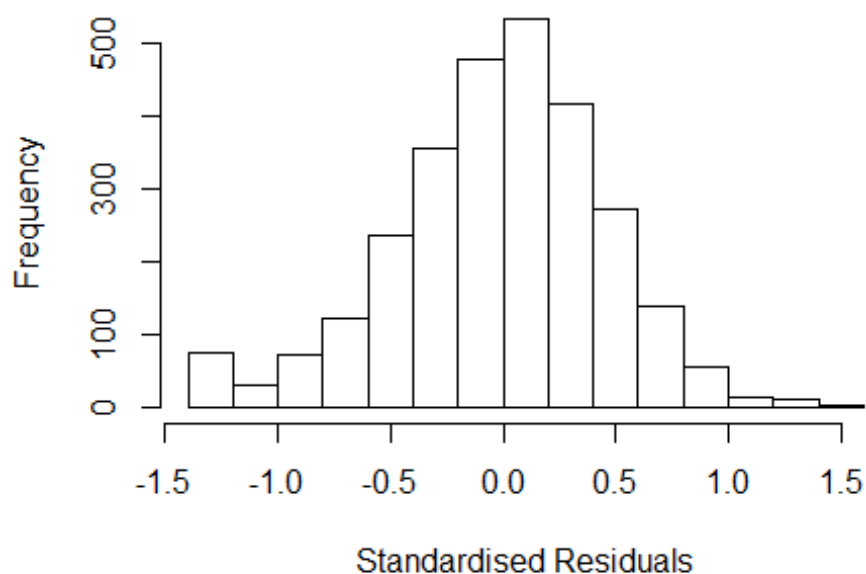
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41803 -0.29789  0.00998  0.29057  1.48594
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2421    0.0698   17.79 < 2e-16 ***
## FirstAuthorFemale1 -0.0015    0.0192   -0.08  0.93736
## LastAuthorFemale1  0.0268    0.0224    1.20  0.23017
## UniqueAuthors2     0.0893    0.0428    2.09  0.03701 *
## UniqueAuthors3     0.1526    0.0406    3.76  0.00018 ***
## UniqueAuthors4     0.2187    0.0422    5.19  2.3e-07 ***
## UniqueAuthors5     0.2630    0.0403    6.53  7.9e-11 ***
## Year1997          -0.0293    0.0870   -0.34  0.73662
## Year1998          -0.0743    0.0793   -0.94  0.34876
## Year1999          -0.0853    0.0739   -1.15  0.24870
```

```

## Year2000          -0.1249      0.0776   -1.61  0.10773
## Year2001          -0.0953      0.0751   -1.27  0.20442
## Year2002          -0.0992      0.0755   -1.31  0.18913
## Year2003          -0.1630      0.0717   -2.27  0.02305 *
## Year2004          -0.1507      0.0742   -2.03  0.04233 *
## Year2005          -0.1491      0.0700   -2.13  0.03325 *
## Year2006          -0.1484      0.0687   -2.16  0.03087 *
## Year2007          -0.1331      0.0671   -1.98  0.04747 *
## Year2008          -0.0712      0.0698   -1.02  0.30804
## Year2009          -0.1124      0.0683   -1.65  0.09984 .
## Year2010          -0.1293      0.0679   -1.90  0.05689 .
## Year2011          -0.1282      0.0686   -1.87  0.06174 .
## Year2012          -0.1106      0.0669   -1.65  0.09855 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0352, Adjusted R-squared:  0.0276
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 236 weights are ~= 1. The remaining 2580 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.222  0.868  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      3.55e-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 1.086 1      1.042
## LastAuthorFemale  1.075 1      1.037
## Year              1.048 16      1.001

```

Residuals from first and last author



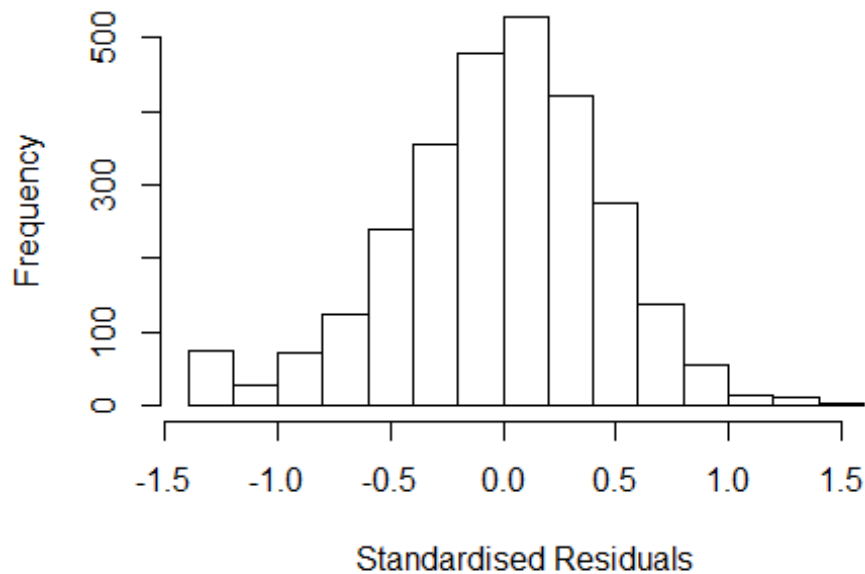
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3752 -0.3111  0.0151  0.2990  1.5433
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3749    0.0617   22.29  <2e-16 ***
## FirstAuthorFemale1  0.0120    0.0193    0.62    0.53
## LastAuthorFemale1  0.0192    0.0225    0.85    0.39
## Year1997          -0.0249    0.0920   -0.27    0.79
## Year1998          -0.0499    0.0792   -0.63    0.53
## Year1999          -0.0746    0.0747   -1.00    0.32
## Year2000          -0.1014    0.0792   -1.28    0.20
## Year2001          -0.0978    0.0763   -1.28    0.20
## Year2002          -0.0606    0.0767   -0.79    0.43
## Year2003          -0.1272    0.0726   -1.75    0.08 .
## Year2004          -0.1198    0.0748   -1.60    0.11
## Year2005          -0.1061    0.0703   -1.51    0.13
```

```

## Year2006          -0.1086      0.0695   -1.56      0.12
## Year2007          -0.0891      0.0681   -1.31      0.19
## Year2008          -0.0309      0.0701   -0.44      0.66
## Year2009          -0.0623      0.0692   -0.90      0.37
## Year2010          -0.0878      0.0688   -1.28      0.20
## Year2011          -0.0808      0.0694   -1.17      0.24
## Year2012          -0.0597      0.0672   -0.89      0.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.00437,    Adjusted R-squared:  -0.00204
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 2561 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  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      3.55e-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 1.03 1      1.015
## Year              1.03 16      1.001

```


Residuals from first author



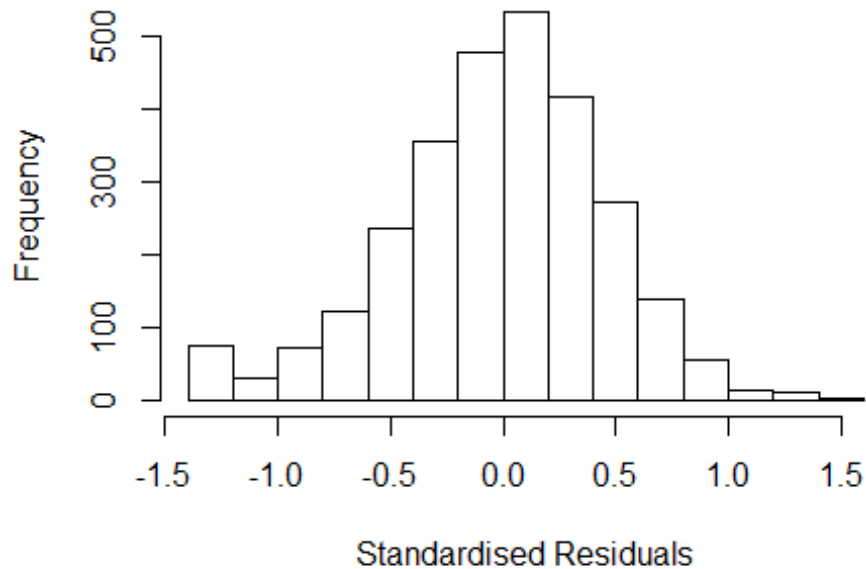
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3636 -0.3030 0.0133 0.2985 1.5407
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3757 0.0617 22.29 <2e-16 ***
## FirstAuthorFemale1 0.0166 0.0188 0.88 0.377
## Year1997 -0.0230 0.0917 -0.25 0.802
## Year1998 -0.0463 0.0791 -0.59 0.558
## Year1999 -0.0732 0.0746 -0.98 0.327
## Year2000 -0.0993 0.0792 -1.25 0.210
## Year2001 -0.0961 0.0762 -1.26 0.208
## Year2002 -0.0578 0.0766 -0.76 0.450
## Year2003 -0.1254 0.0725 -1.73 0.084 .
## Year2004 -0.1172 0.0747 -1.57 0.117
## Year2005 -0.1051 0.0703 -1.49 0.135
## Year2006 -0.1064 0.0694 -1.53 0.125
```

```

## Year2007          -0.0879      0.0680    -1.29     0.197
## Year2008          -0.0288      0.0700    -0.41     0.681
## Year2009          -0.0607      0.0691    -0.88     0.380
## Year2010          -0.0852      0.0686    -1.24     0.215
## Year2011          -0.0788      0.0692    -1.14     0.255
## Year2012          -0.0585      0.0672    -0.87     0.384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.00408,    Adjusted R-squared:  -0.00197
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 249 weights are ~= 1. The remaining 2567 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  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      3.55e-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 1.02 1          1.010
## Year              1.02 16          1.001

```

Residuals from last author



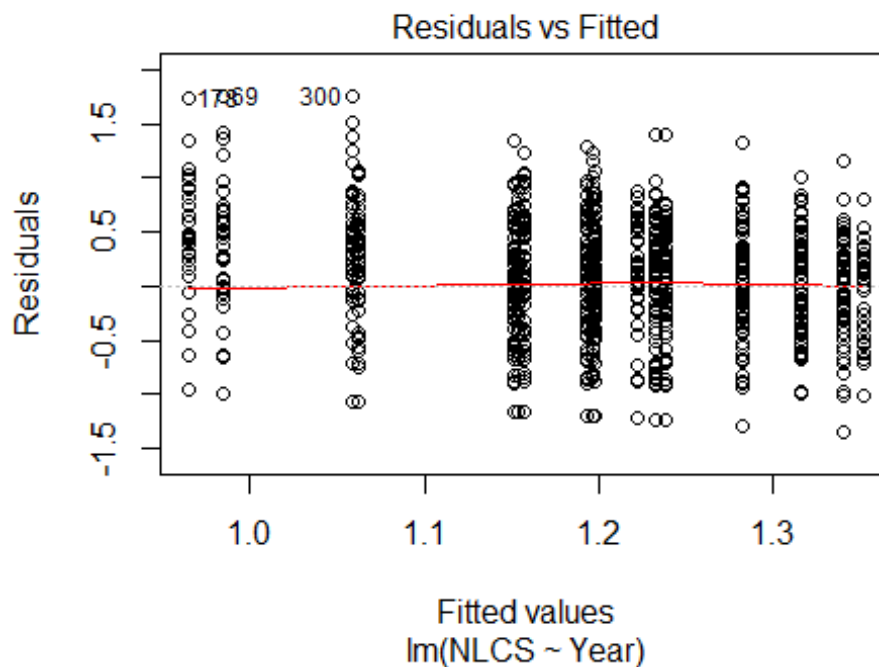
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3696 -0.3087 0.0151 0.2973 1.5408
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3768 0.0618 22.28 <2e-16 ***
## LastAuthorFemale1 0.0230 0.0220 1.04 0.296
## Year1997 -0.0248 0.0920 -0.27 0.787
## Year1998 -0.0504 0.0793 -0.64 0.525
## Year1999 -0.0739 0.0748 -0.99 0.323
## Year2000 -0.1007 0.0793 -1.27 0.204
## Year2001 -0.0966 0.0762 -1.27 0.205
## Year2002 -0.0601 0.0768 -0.78 0.434
## Year2003 -0.1266 0.0726 -1.74 0.082 .
## Year2004 -0.1178 0.0747 -1.58 0.115
## Year2005 -0.1047 0.0703 -1.49 0.137
## Year2006 -0.1068 0.0694 -1.54 0.124
```

```

## Year2007          -0.0877      0.0681    -1.29      0.197
## Year2008          -0.0302      0.0702    -0.43      0.668
## Year2009          -0.0609      0.0691    -0.88      0.378
## Year2010          -0.0864      0.0687    -1.26      0.209
## Year2011          -0.0795      0.0694    -1.15      0.252
## Year2012          -0.0577      0.0671    -0.86      0.390
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.00422,    Adjusted R-squared:  -0.00183
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 253 weights are ~= 1. The remaining 2563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  0.867  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      3.55e-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: 2816"
## [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
##   98   92   95   72   85   75   70   67   93  106   89  100   86  114   95
## 2011 2012
##  113  107
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   50   52   61   44   36   40   53   51   60   81   66   71   69   86   70
## 2011 2012

```

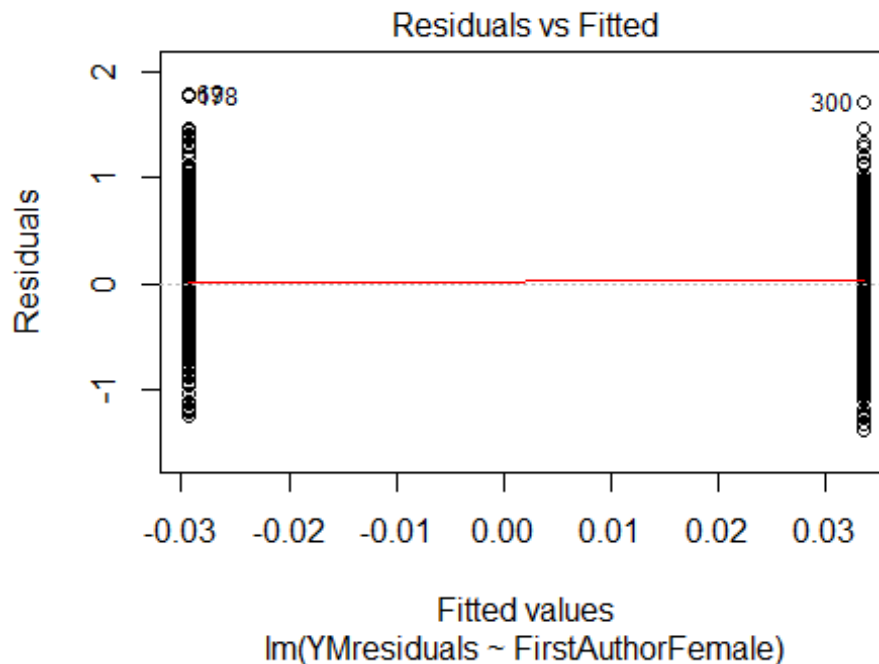
```
##      84      78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    47   50   57   42   31   36   50   46   56   75   61   64   63   80   64
## 2011 2012
##    78   63
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett 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 = 7.6, df = 1, p-value = 0.006
## [1] "Female first author team size 2018 geometric mean: 3.64379765029341"
## [1] "Male first author team size 2018 geometric mean: 4.13540241364536"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

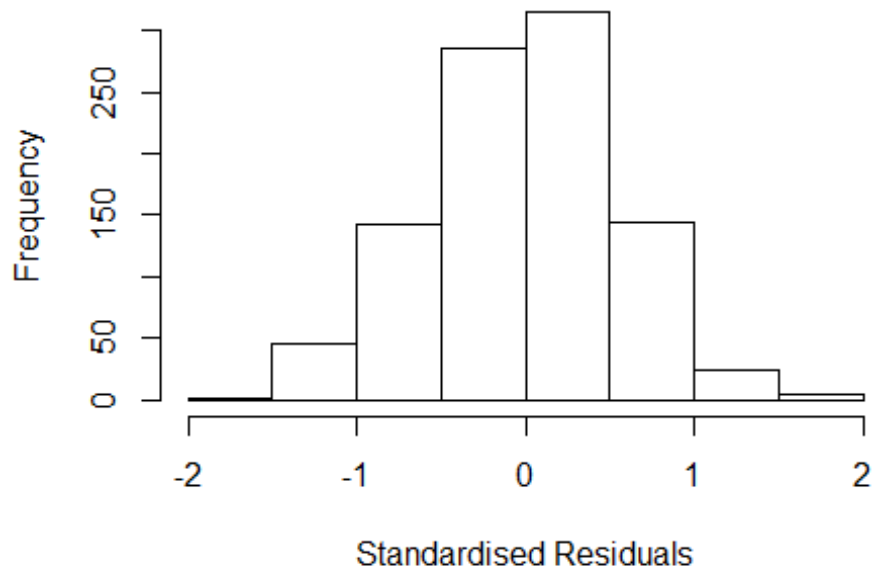
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 710, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.34161721611295"
## [1] "Male last author team size 2018 geometric mean: 4.25121753617776"

## 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 = 610, 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 1.429  1      1.196
## LastAuthorFemale  1.539  1      1.241
## UniqueAuthors    1.334  4      1.037
## Year              1.334 16      1.009
```

Residuals from first and last author and team size



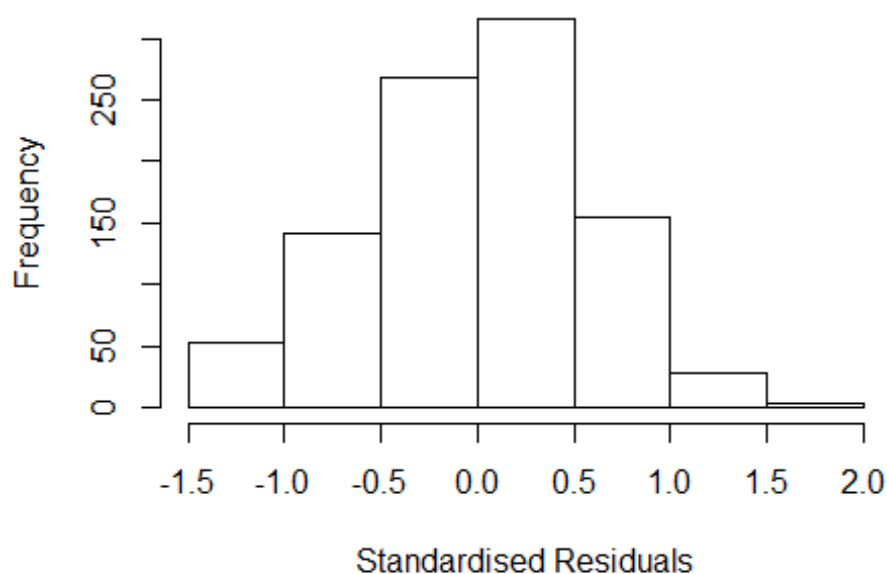
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5052 -0.3753 0.0232 0.4000 1.8578
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7451 0.1477 5.04 5.5e-07 ***
## FirstAuthorFemale1 0.0417 0.0440 0.95 0.3441
## LastAuthorFemale1 0.0615 0.0464 1.32 0.1857
## UniqueAuthors2 0.1958 0.0758 2.58 0.0099 **
## UniqueAuthors3 0.2331 0.0742 3.14 0.0017 **
## UniqueAuthors4 0.3895 0.0756 5.16 3.1e-07 ***
## UniqueAuthors5 0.4012 0.0736 5.45 6.4e-08 ***
## Year1997 -0.0917 0.1824 -0.50 0.6155
## Year1998 0.0346 0.1809 0.19 0.8482
## Year1999 0.2413 0.1669 1.45 0.1486
```

```

## Year2000          0.3018      0.1528      1.98      0.0486 *
## Year2001          0.1822      0.1557      1.17      0.2423
## Year2002          0.1944      0.1558      1.25      0.2125
## Year2003          0.0677      0.1712      0.40      0.6927
## Year2004          0.1118      0.1609      0.69      0.4873
## Year2005          0.1621      0.1479      1.10      0.2733
## Year2006          0.0993      0.1494      0.66      0.5065
## Year2007          0.1558      0.1527      1.02      0.3079
## Year2008          0.2674      0.1431      1.87      0.0620 .
## Year2009          0.2751      0.1400      1.97      0.0497 *
## Year2010          0.2166      0.1492      1.45      0.1469
## Year2011          0.1240      0.1466      0.85      0.3978
## Year2012          0.2610      0.1478      1.77      0.0776 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.092, Adjusted R-squared:  0.0708
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 893 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  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      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 1.365 1      1.168
## LastAuthorFemale  1.386 1      1.177
## Year              1.147 16      1.004

```


Residuals from first and last author



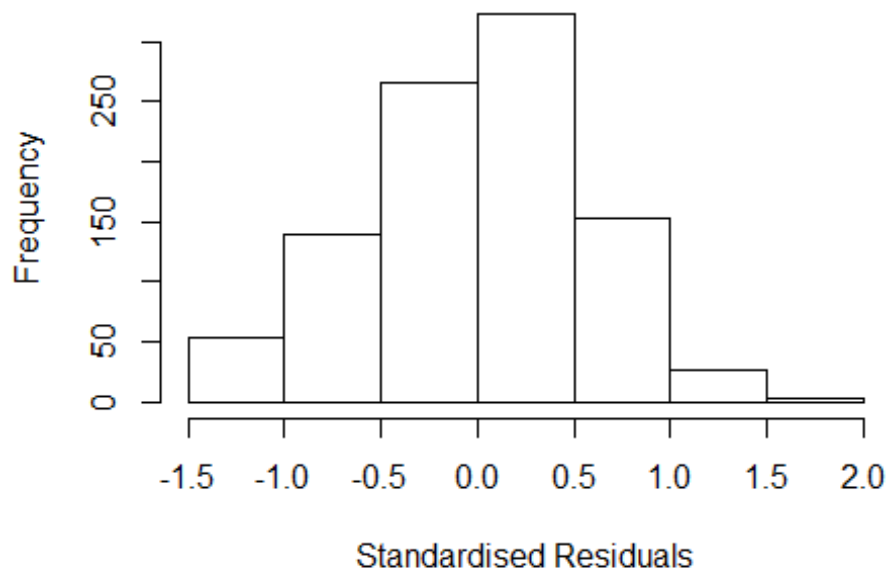
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3839 -0.3761 0.0215 0.3883 1.8481
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9674 0.1500 6.45 1.8e-10 ***
## FirstAuthorFemale1 0.0662 0.0439 1.51 0.132
## LastAuthorFemale1 0.0173 0.0448 0.39 0.700
## Year1997 -0.1084 0.1965 -0.55 0.581
## Year1998 0.0621 0.1928 0.32 0.747
## Year1999 0.2572 0.1822 1.41 0.158
## Year2000 0.3251 0.1678 1.94 0.053 .
## Year2001 0.2045 0.1710 1.20 0.232
## Year2002 0.2303 0.1692 1.36 0.174
## Year2003 0.0744 0.1821 0.41 0.683
## Year2004 0.1687 0.1754 0.96 0.336
## Year2005 0.2078 0.1625 1.28 0.201
```

```

## Year2006          0.1617      0.1639      0.99      0.324
## Year2007          0.2308      0.1660      1.39      0.165
## Year2008          0.3331      0.1586      2.10      0.036 *
## Year2009          0.3151      0.1550      2.03      0.042 *
## Year2010          0.2761      0.1639      1.68      0.092 .
## Year2011          0.2061      0.1604      1.29      0.199
## Year2012          0.3500      0.1609      2.18      0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0461, Adjusted R-squared:  0.0279
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 872 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.246  0.861  0.948  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 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



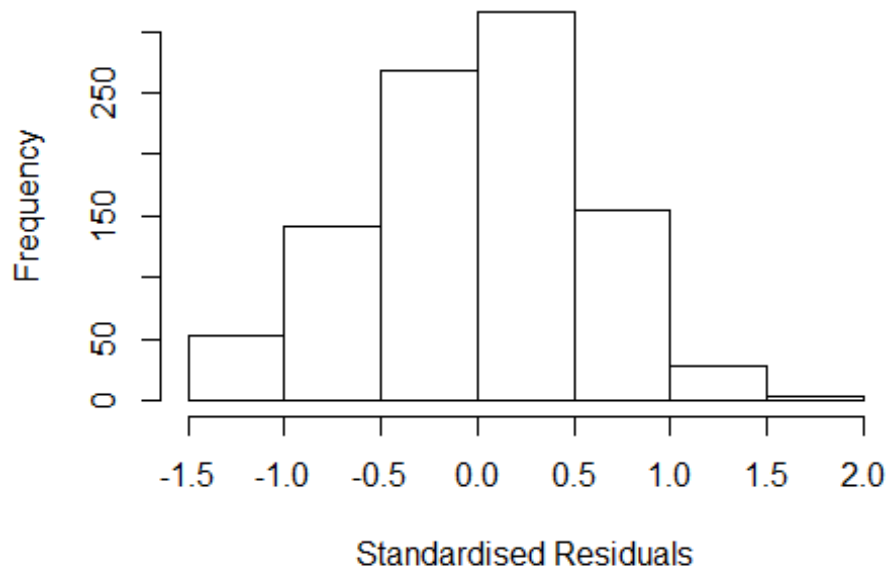
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3786 -0.3701 0.0254 0.3860 1.8466
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9700 0.1492 6.50 1.3e-10 ***
## FirstAuthorFemale1 0.0740 0.0389 1.90 0.058 .
## Year1997 -0.1096 0.1975 -0.55 0.579
## Year1998 0.0590 0.1928 0.31 0.760
## Year1999 0.2576 0.1804 1.43 0.154
## Year2000 0.3248 0.1676 1.94 0.053 .
## Year2001 0.2052 0.1710 1.20 0.230
## Year2002 0.2296 0.1682 1.37 0.173
## Year2003 0.0771 0.1817 0.42 0.671
## Year2004 0.1693 0.1751 0.97 0.334
## Year2005 0.2070 0.1618 1.28 0.201
## Year2006 0.1607 0.1638 0.98 0.327
```

```

## Year2007          0.2315      0.1658      1.40      0.163
## Year2008          0.3346      0.1587      2.11      0.035 *
## Year2009          0.3139      0.1552      2.02      0.043 *
## Year2010          0.2761      0.1633      1.69      0.091 .
## Year2011          0.2053      0.1601      1.28      0.200
## Year2012          0.3486      0.1606      2.17      0.030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0463, Adjusted R-squared:  0.0291
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 879 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.235  0.859  0.948  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      1.04e-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 1.085 1      1.041
## Year              1.085 16      1.003

```

Residuals from last author



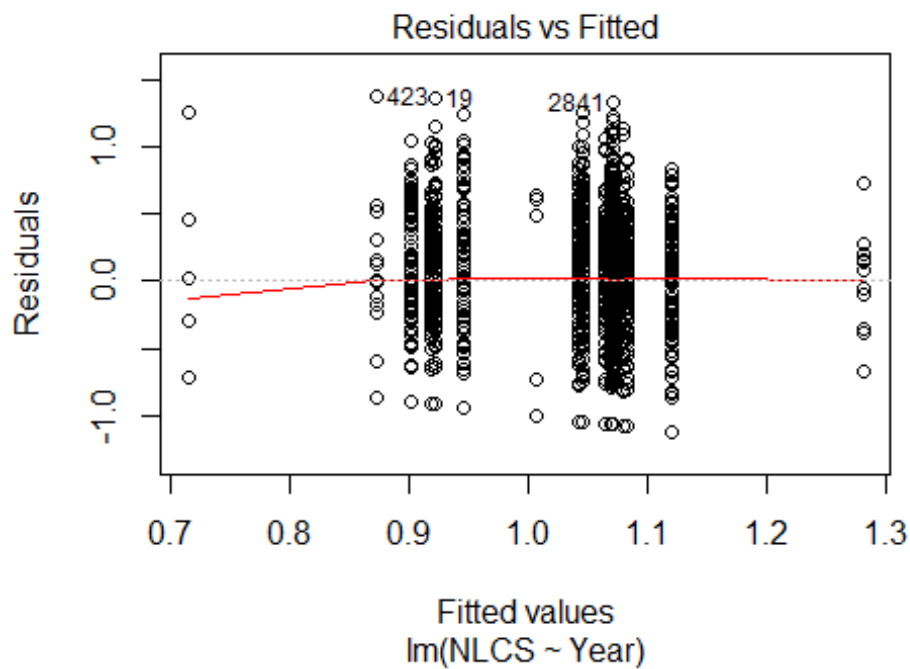
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3761 -0.3646 0.0315 0.3810 1.8343
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9776 0.1529 6.39 2.5e-10 ***
## LastAuthorFemale1 0.0490 0.0398 1.23 0.218
## Year1997 -0.1050 0.2010 -0.52 0.602
## Year1998 0.0669 0.1967 0.34 0.734
## Year1999 0.2683 0.1827 1.47 0.142
## Year2000 0.3381 0.1697 1.99 0.047 *
## Year2001 0.2079 0.1733 1.20 0.231
## Year2002 0.2361 0.1721 1.37 0.170
## Year2003 0.0853 0.1839 0.46 0.643
## Year2004 0.1779 0.1774 1.00 0.316
## Year2005 0.2199 0.1644 1.34 0.181
## Year2006 0.1748 0.1668 1.05 0.295
```

```

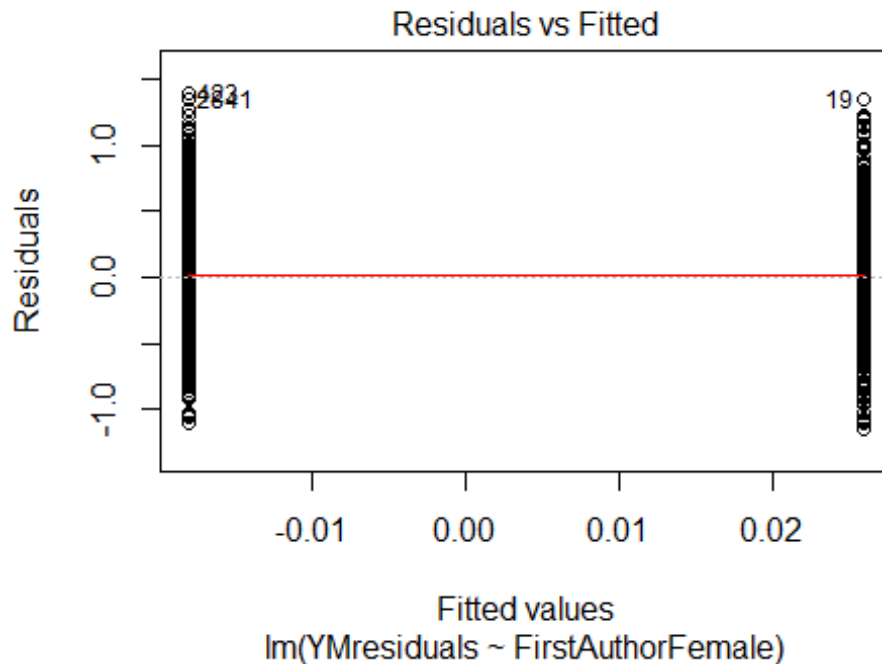
## Year2007          0.2380      0.1686      1.41      0.158
## Year2008          0.3494      0.1608      2.17      0.030 *
## Year2009          0.3338      0.1579      2.11      0.035 *
## Year2010          0.2918      0.1656      1.76      0.078 .
## Year2011          0.2216      0.1632      1.36      0.175
## Year2012          0.3641      0.1633      2.23      0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0439, Adjusted R-squared:  0.0267
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 878 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.244  0.857   0.945   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      1.04e-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: 963"
## [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
## 181 186 9 19 14 203 16 150 141 134 166 161 166 207 194
## 2011 2012
## 260 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 126 5 15 11 122 6 115 104 89 119 115 134 158 162
## 2011 2012

```

```
## 201 214
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 103 119 4 15 9 111 6 103 89 81 110 103 117 147 151
## 2011 2012
## 188 197
## [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 = 0.2
```

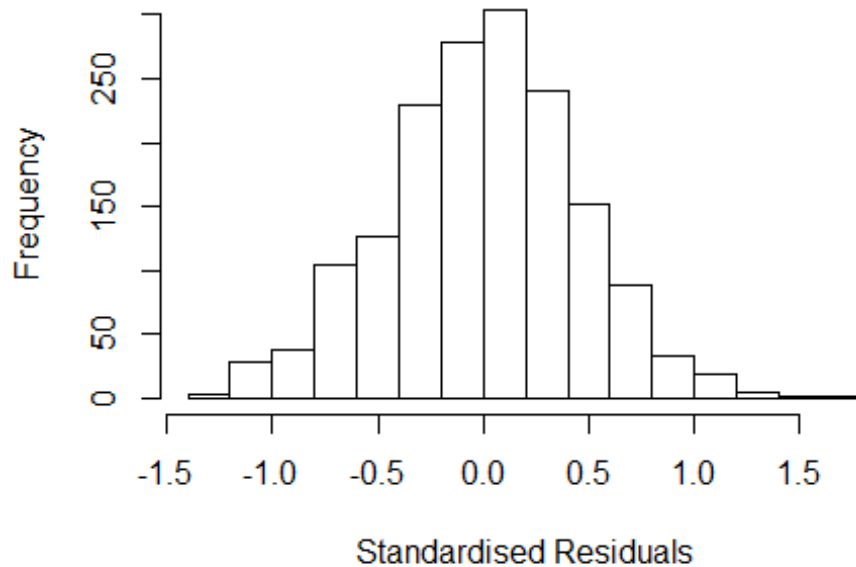


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.86, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.56664255554896"
## [1] "Male first author team size 2018 geometric mean: 3.06248981426626"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7400, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.29877547743376"
## [1] "Male last author team size 2018 geometric mean: 3.33680183914744"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.097 1          1.047
## LastAuthorFemale  1.112 1          1.055
## UniqueAuthors    1.208 4          1.024
## Year             1.252 16          1.007
```


Residuals from first and last author and team size



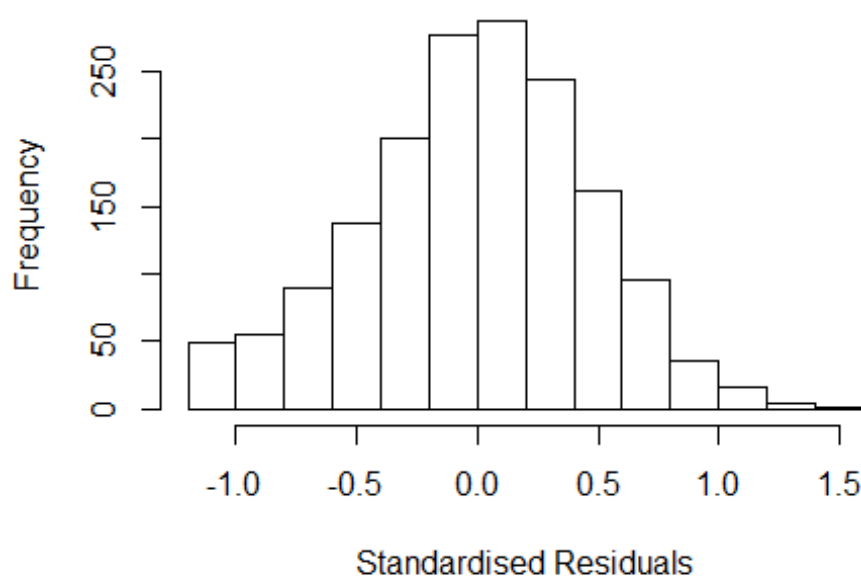
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2589 -0.2897 0.0097 0.2915 1.6189
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7015 0.0518 13.54 < 2e-16 ***
## FirstAuthorFemale1 0.0267 0.0236 1.13 0.2580
## LastAuthorFemale1 0.0026 0.0259 0.10 0.9200
## UniqueAuthors2 0.2701 0.0358 7.54 7.8e-14 ***
## UniqueAuthors3 0.3174 0.0367 8.65 < 2e-16 ***
## UniqueAuthors4 0.3196 0.0413 7.74 1.8e-14 ***
## UniqueAuthors5 0.4207 0.0368 11.42 < 2e-16 ***
## Year1997 0.0239 0.0697 0.34 0.7312
## Year1998 -0.0634 0.5183 -0.12 0.9027
## Year1999 -0.0488 0.1356 -0.36 0.7188
```

```

## Year2000          0.3289      0.1115      2.95      0.0032 **
## Year2001         -0.0151      0.0622     -0.24      0.8081
## Year2002         -0.3504      0.2775     -1.26      0.2068
## Year2003          0.0656      0.0663      0.99      0.3227
## Year2004         -0.0699      0.0727     -0.96      0.3364
## Year2005          0.1086      0.0672      1.61      0.1065
## Year2006          0.1100      0.0618      1.78      0.0755 .
## Year2007          0.1267      0.0648      1.95      0.0509 .
## Year2008          0.0877      0.0630      1.39      0.1644
## Year2009          0.1095      0.0578      1.90      0.0581 .
## Year2010          0.0857      0.0594      1.44      0.1493
## Year2011          0.0763      0.0583      1.31      0.1909
## Year2012          0.0912      0.0575      1.59      0.1131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.124, Adjusted R-squared:  0.113
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1508 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.869  0.950  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      6.05e-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 1.051 1      1.025
## LastAuthorFemale  1.074 1      1.036
## Year              1.080 16      1.002

```

Residuals from first and last author



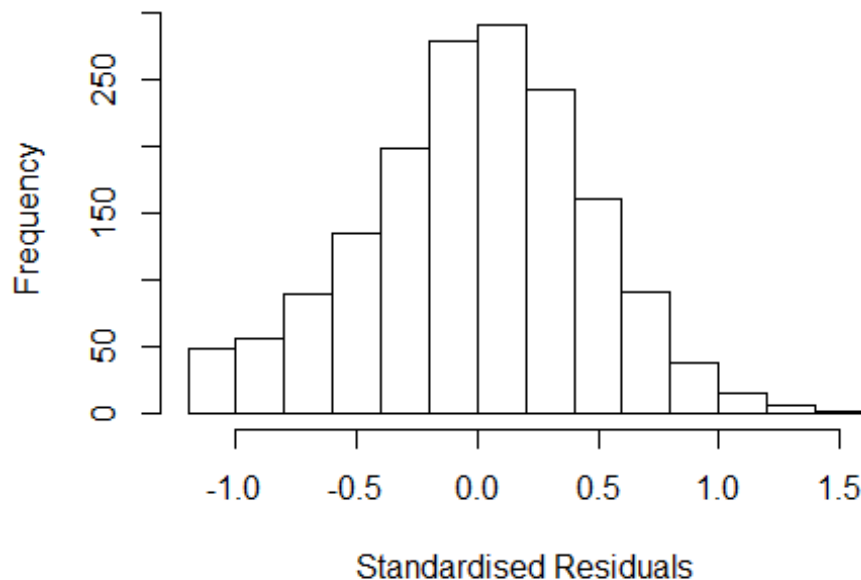
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1711 -0.3048  0.0114  0.3211  1.4294
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9009    0.0500   18.03  <2e-16 ***
## FirstAuthorFemale1  0.0647    0.0243    2.66  0.0079 **
## LastAuthorFemale1 -0.0255    0.0266   -0.96  0.3368
## Year1997         0.0320    0.0736    0.44  0.6633
## Year1998        -0.0694    0.6898   -0.10  0.9199
## Year1999        -0.0813    0.1420   -0.57  0.5673
## Year2000         0.3943    0.1318    2.99  0.0028 **
## Year2001         0.0234    0.0681    0.34  0.7317
## Year2002        -0.3424    0.3264   -1.05  0.2943
## Year2003         0.1194    0.0702    1.70  0.0892 .
## Year2004        -0.0270    0.0790   -0.34  0.7326
## Year2005         0.1740    0.0704    2.47  0.0135 *
```

```

## Year2006          0.1643      0.0642      2.56      0.0106 *
## Year2007          0.2055      0.0656      3.13      0.0018 **
## Year2008          0.1507      0.0646      2.33      0.0199 *
## Year2009          0.1626      0.0610      2.66      0.0078 **
## Year2010          0.1377      0.0616      2.23      0.0256 *
## Year2011          0.1378      0.0610      2.26      0.0240 *
## Year2012          0.1541      0.0599      2.57      0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.0202
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1510 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.315  0.865   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      6.05e-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 1.033 1      1.016
## Year              1.033 16      1.001

```

Residuals from first author



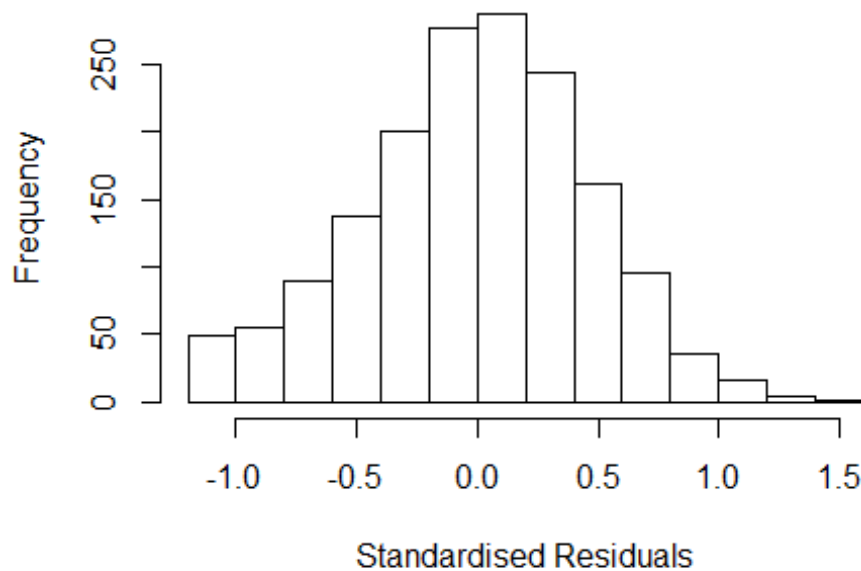
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1599 -0.3124 0.0105 0.3232 1.4315
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8956 0.0498 18.00 <2e-16 ***
## FirstAuthorFemale1 0.0590 0.0243 2.43 0.0152 *
## Year1997 0.0339 0.0738 0.46 0.6459
## Year1998 -0.0655 0.6760 -0.10 0.9229
## Year1999 -0.0781 0.1427 -0.55 0.5842
## Year2000 0.3895 0.1326 2.94 0.0034 **
## Year2001 0.0235 0.0684 0.34 0.7314
## Year2002 -0.3437 0.3235 -1.06 0.2882
## Year2003 0.1180 0.0702 1.68 0.0931 .
## Year2004 -0.0261 0.0792 -0.33 0.7418
## Year2005 0.1758 0.0707 2.49 0.0130 *
## Year2006 0.1653 0.0643 2.57 0.0103 *
```

```

## Year2007          0.2054      0.0659      3.12      0.0019 **
## Year2008          0.1492      0.0646      2.31      0.0210 *
## Year2009          0.1615      0.0612      2.64      0.0084 **
## Year2010          0.1383      0.0618      2.24      0.0254 *
## Year2011          0.1390      0.0611      2.27      0.0232 *
## Year2012          0.1550      0.0600      2.58      0.0099 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.0203
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.313  0.866  0.950  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      6.05e-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 1.053 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
## -1.1258 -0.3036  0.0202  0.3145  1.4142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91946    0.04941  18.61  <2e-16 ***
## LastAuthorFemale1 -0.00813    0.02641  -0.31  0.7582
## Year1997         0.03132    0.07379   0.42  0.6713
## Year1998        -0.05103    0.72460  -0.07  0.9439
## Year1999        -0.08470    0.14439  -0.59  0.5575
## Year2000         0.39449    0.13498   2.92  0.0035 **
## Year2001         0.02113    0.06838   0.31  0.7573
## Year2002        -0.34570    0.33371  -1.04  0.3004
## Year2003         0.12382    0.06993   1.77  0.0768 .
## Year2004        -0.02533    0.07943  -0.32  0.7498
## Year2005         0.17913    0.07127   2.51  0.0121 *
## Year2006         0.17190    0.06390   2.69  0.0072 **
```

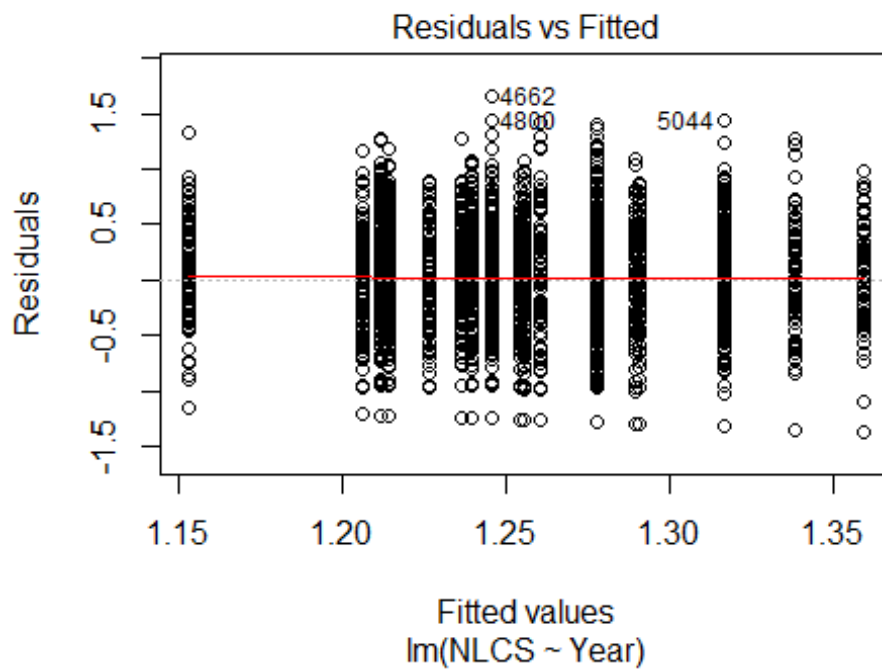
```

## Year2007      0.20639      0.06613      3.12      0.0018 **
## Year2008      0.15862      0.06437      2.46      0.0138 *
## Year2009      0.16526      0.06120      2.70      0.0070 **
## Year2010      0.14239      0.06178      2.30      0.0213 *
## Year2011      0.14234      0.06121      2.33      0.0202 *
## Year2012      0.15833      0.05976      2.65      0.0081 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.0165
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 153 weights are ~= 1. The remaining 1500 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.327  0.861  0.948  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      6.05e-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: 1653"
## [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
## 196 232 265 241 267 254 256 245 249 267 287 316 309 363 345
## 2011 2012
## 378 402
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 73 97 108 106 92 82 105 109 128 170 205 242 246 286 269
## 2011 2012

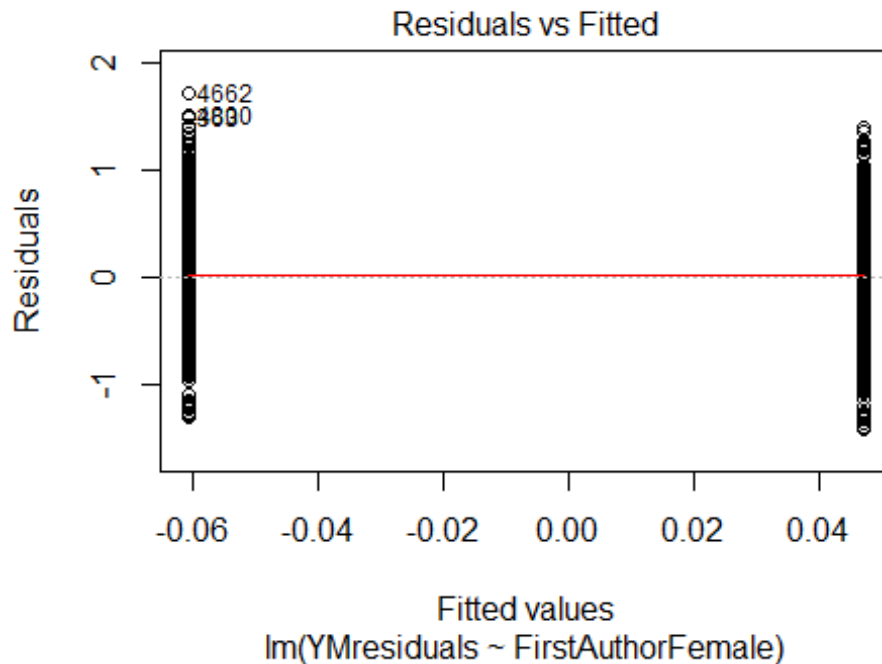
```



```
## 294 330
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 89 96 94 84 68 89 91 108 144 187 222 223 254 244
## 2011 2012
## 272 307
## [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
```

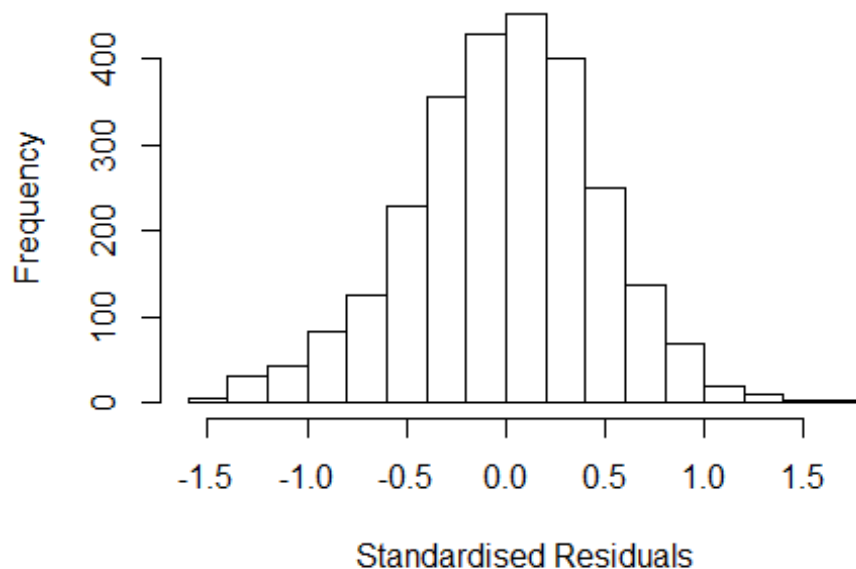


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 24, df = 1, p-value = 8e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.40543790792537"
## [1] "Male first author team size 2018 geometric mean: 3.33864391344129"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, p-value = 0.004
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.32383114037217"
## [1] "Male last author team size 2018 geometric mean: 3.67213658924916"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21000, 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 1.120 1      1.058
## LastAuthorFemale  1.119 1      1.058
## UniqueAuthors    1.193 4      1.022
## Year              1.293 16      1.008
```

Residuals from first and last author and team size



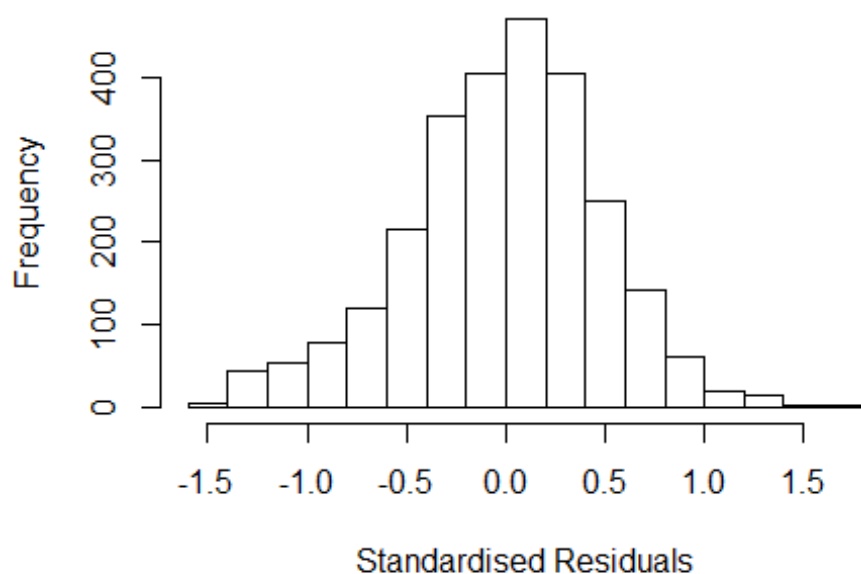
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5035 -0.3015 0.0121 0.3031 1.7370
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0678 0.0754 14.16 < 2e-16 ***
## FirstAuthorFemale1 0.0722 0.0201 3.59 0.00034 ***
## LastAuthorFemale1 0.0872 0.0197 4.42 1.0e-05 ***
## UniqueAuthors2 0.2435 0.0475 5.13 3.1e-07 ***
## UniqueAuthors3 0.2698 0.0454 5.94 3.3e-09 ***
## UniqueAuthors4 0.3245 0.0464 6.99 3.4e-12 ***
## UniqueAuthors5 0.4193 0.0440 9.54 < 2e-16 ***
## Year1997 -0.1397 0.0956 -1.46 0.14413
## Year1998 -0.0856 0.0883 -0.97 0.33249
## Year1999 -0.2488 0.0825 -3.02 0.00258 **
```

```

## Year2000          -0.1126      0.0855    -1.32   0.18829
## Year2001          -0.0767      0.0915    -0.84   0.40159
## Year2002          -0.1429      0.0860    -1.66   0.09672 .
## Year2003          -0.1983      0.0836    -2.37   0.01782 *
## Year2004          -0.2170      0.0814    -2.67   0.00773 **
## Year2005          -0.2171      0.0777    -2.80   0.00522 **
## Year2006          -0.1889      0.0781    -2.42   0.01565 *
## Year2007          -0.2071      0.0770    -2.69   0.00717 **
## Year2008          -0.2524      0.0761    -3.32   0.00093 ***
## Year2009          -0.2306      0.0771    -2.99   0.00281 **
## Year2010          -0.2312      0.0771    -3.00   0.00275 **
## Year2011          -0.1304      0.0763    -1.71   0.08725 .
## Year2012          -0.1977      0.0756    -2.61   0.00899 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0859, Adjusted R-squared:  0.0782
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.117  0.871  0.953   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          3.79e-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 1.107 1          1.052
## LastAuthorFemale  1.103 1          1.050
## Year              1.097 16          1.003

```

Residuals from first and last author



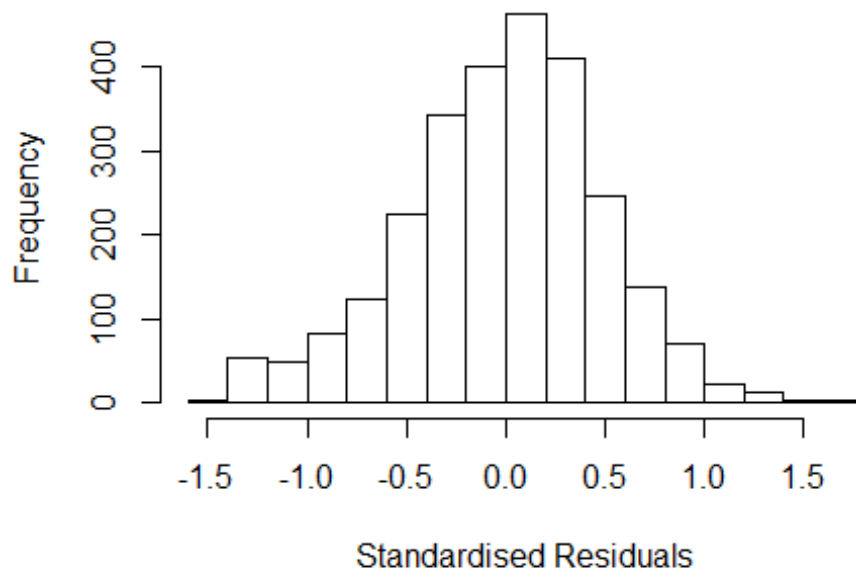
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4591 -0.3106  0.0202  0.3031  1.7492
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2912     0.0691   18.70 < 2e-16 ***
## FirstAuthorFemale1  0.0984     0.0206    4.78 1.8e-06 ***
## LastAuthorFemale1  0.0824     0.0200    4.12 3.9e-05 ***
## Year1997          -0.1089     0.0967   -1.13  0.260
## Year1998          -0.0447     0.0874   -0.51  0.609
## Year1999          -0.1966     0.0828   -2.37  0.018 *
## Year2000          -0.0774     0.0878   -0.88  0.378
## Year2001          -0.0129     0.0880   -0.15  0.884
## Year2002          -0.0775     0.0824   -0.94  0.347
## Year2003          -0.1368     0.0826   -1.66  0.098 .
## Year2004          -0.1527     0.0811   -1.88  0.060 .
## Year2005          -0.1375     0.0765   -1.80  0.072 .
```

```

## Year2006          -0.1088      0.0771   -1.41    0.159
## Year2007          -0.1346      0.0759   -1.77    0.076 .
## Year2008          -0.1749      0.0757   -2.31    0.021 *
## Year2009          -0.1565      0.0762   -2.05    0.040 *
## Year2010          -0.1423      0.0758   -1.88    0.060 .
## Year2011          -0.0503      0.0748   -0.67    0.501
## Year2012          -0.1080      0.0745   -1.45    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.0237
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 207 weights are ~= 1. The remaining 2432 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.125  0.870   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      3.79e-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 1.047 1      1.023
## Year              1.047 16      1.001

```

Residuals from first author



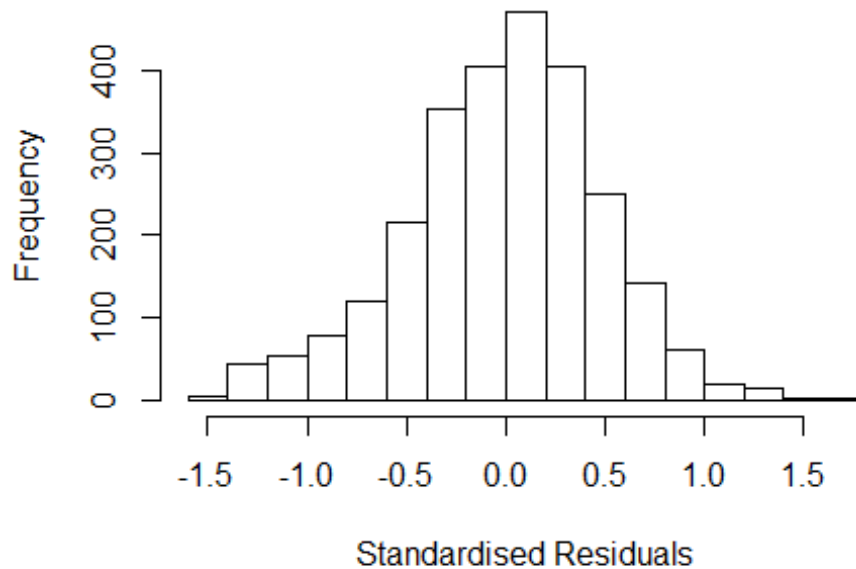
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4275 -0.3154 0.0173 0.3104 1.7246
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30922 0.06793 19.27 < 2e-16 ***
## FirstAuthorFemale1 0.11829 0.02011 5.88 4.6e-09 ***
## Year1997 -0.10766 0.09603 -1.12 0.262
## Year1998 -0.05102 0.08634 -0.59 0.555
## Year1999 -0.19343 0.08154 -2.37 0.018 *
## Year2000 -0.07146 0.08684 -0.82 0.411
## Year2001 -0.00974 0.08814 -0.11 0.912
## Year2002 -0.07824 0.08155 -0.96 0.337
## Year2003 -0.14008 0.08186 -1.71 0.087 .
## Year2004 -0.15744 0.07993 -1.97 0.049 *
## Year2005 -0.13728 0.07580 -1.81 0.070 .
## Year2006 -0.09781 0.07649 -1.28 0.201
```

```

## Year2007          -0.12707    0.07509   -1.69    0.091 .
## Year2008          -0.16546    0.07518   -2.20    0.028 *
## Year2009          -0.15188    0.07548   -2.01    0.044 *
## Year2010          -0.13583    0.07515   -1.81    0.071 .
## Year2011          -0.03599    0.07390   -0.49    0.626
## Year2012          -0.09794    0.07363   -1.33    0.184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.024, Adjusted R-squared:  0.0176
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 216 weights are ~= 1. The remaining 2423 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.140  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      3.79e-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 1.049 1          1.024
## Year            1.049 16          1.002

```


Residuals from last author



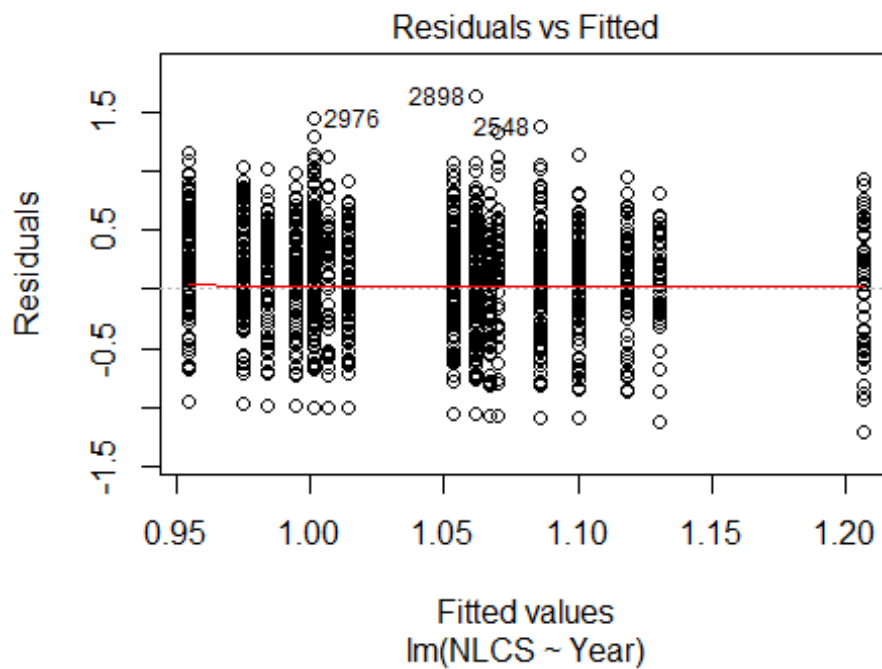
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4248 -0.3146  0.0192  0.3117  1.6954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.32290    0.07037   18.80 < 2e-16 ***
## LastAuthorFemale1 0.10595    0.01957    5.41 6.7e-08 ***
## Year1997       -0.11189    0.09764   -1.15  0.252
## Year1998       -0.03052    0.08914   -0.34  0.732
## Year1999       -0.18998    0.08408   -2.26  0.024 *
## Year2000       -0.06691    0.08952   -0.75  0.455
## Year2001       -0.00409    0.09040   -0.05  0.964
## Year2002       -0.06108    0.08395   -0.73  0.467
## Year2003       -0.12779    0.08349   -1.53  0.126
## Year2004       -0.13868    0.08212   -1.69  0.091 .
## Year2005       -0.12480    0.07775   -1.61  0.109
## Year2006       -0.10056    0.07862   -1.28  0.201
```

```

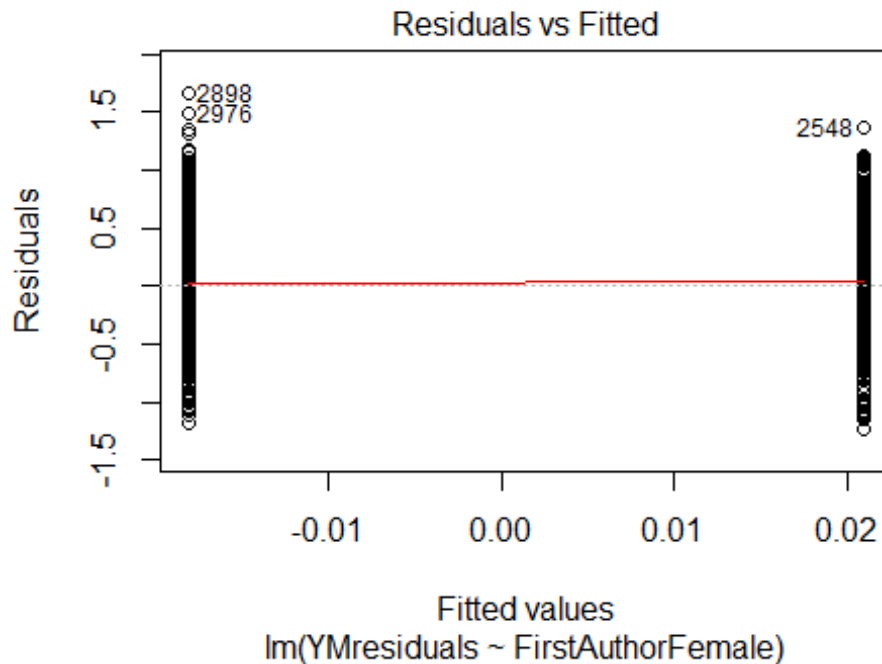
## Year2007          -0.11403      0.07702    -1.48      0.139
## Year2008          -0.15660      0.07695    -2.04      0.042 *
## Year2009          -0.13655      0.07735    -1.77      0.078 .
## Year2010          -0.12030      0.07701    -1.56      0.118
## Year2011          -0.03247      0.07607    -0.43      0.669
## Year2012          -0.08861      0.07567    -1.17      0.242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.0208, Adjusted R-squared:  0.0144
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 194 weights are ~= 1. The remaining 2445 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.159  0.870  0.954  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      3.79e-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: 2639"
## [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
##   77   85   85   97  124  123  111  109  127  143  171  141  212  213  219
## 2011 2012
##  263  247
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   46   54   62   73   48   90   83   93  102  125  103  162  154  161
## 2011 2012

```

```
## 187 180
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 38 49 54 69 44 83 76 78 92 109 90 151 140 144
## 2011 2012
## 169 168
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

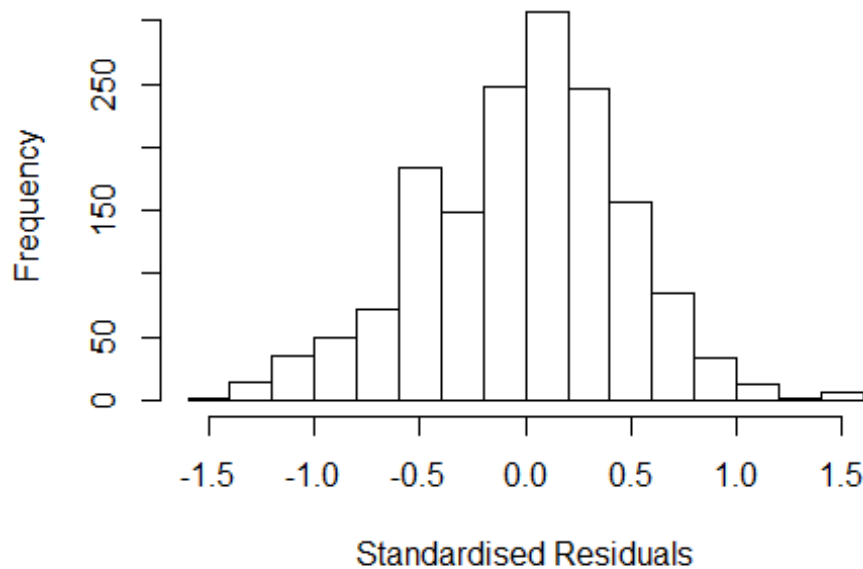


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.9, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.36741496494666"
## [1] "Male first author team size 2018 geometric mean: 3.70621488074199"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7200, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.82327299966939"
## [1] "Male last author team size 2018 geometric mean: 4.31146623960051"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5900, 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 1.131 1      1.064
## LastAuthorFemale  1.108 1      1.053
## UniqueAuthors    1.237 4      1.027
## Year             1.337 16      1.009
```

Residuals from first and last author and team size



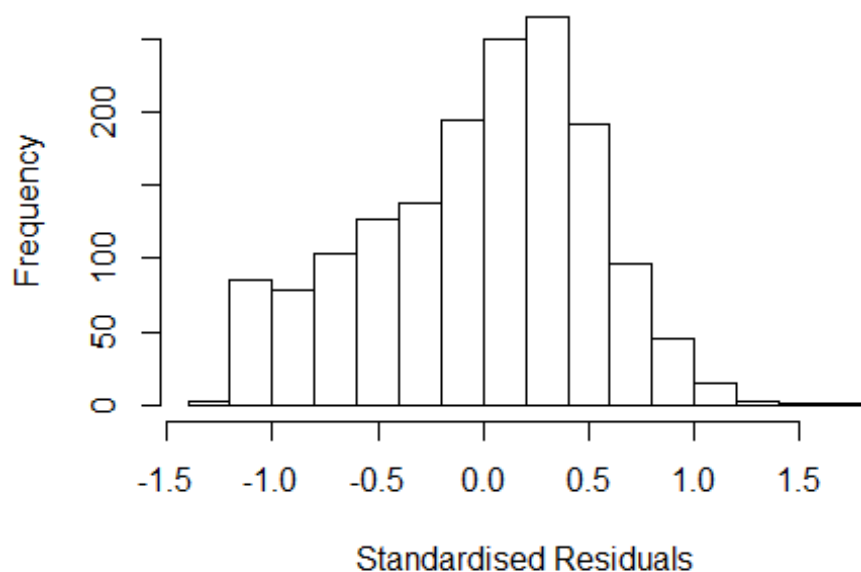
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4081 -0.3322 0.0413 0.3063 1.5798
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5870 0.0822 7.14 1.4e-12 ***
## FirstAuthorFemale1 0.0190 0.0258 0.74 0.46
## LastAuthorFemale1 -0.0238 0.0273 -0.87 0.38
## UniqueAuthors2 0.4016 0.0485 8.28 2.7e-16 ***
## UniqueAuthors3 0.5581 0.0472 11.84 < 2e-16 ***
## UniqueAuthors4 0.6467 0.0438 14.76 < 2e-16 ***
## UniqueAuthors5 0.7038 0.0385 18.30 < 2e-16 ***
## Year1997 0.0522 0.1090 0.48 0.63
## Year1998 0.1831 0.1176 1.56 0.12
## Year1999 0.0103 0.1089 0.09 0.92
```

```

## Year2000          0.0650      0.0955      0.68      0.50
## Year2001          0.1173      0.1025      1.14      0.25
## Year2002          0.0530      0.0925      0.57      0.57
## Year2003         -0.0910      0.0924     -0.98      0.32
## Year2004         -0.0449      0.0948     -0.47      0.64
## Year2005         -0.0480      0.0863     -0.56      0.58
## Year2006         -0.0670      0.0850     -0.79      0.43
## Year2007         -0.1069      0.0930     -1.15      0.25
## Year2008         -0.0944      0.0892     -1.06      0.29
## Year2009         -0.0949      0.0887     -1.07      0.29
## Year2010         -0.0207      0.0854     -0.24      0.81
## Year2011         -0.0228      0.0864     -0.26      0.79
## Year2012         -0.0685      0.0866     -0.79      0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.234, Adjusted R-squared:  0.224
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1477 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.867  0.947  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      6.25e-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 1.069 1      1.034
## LastAuthorFemale  1.043 1      1.021
## Year              1.095 16      1.003

```

Residuals from first and last author



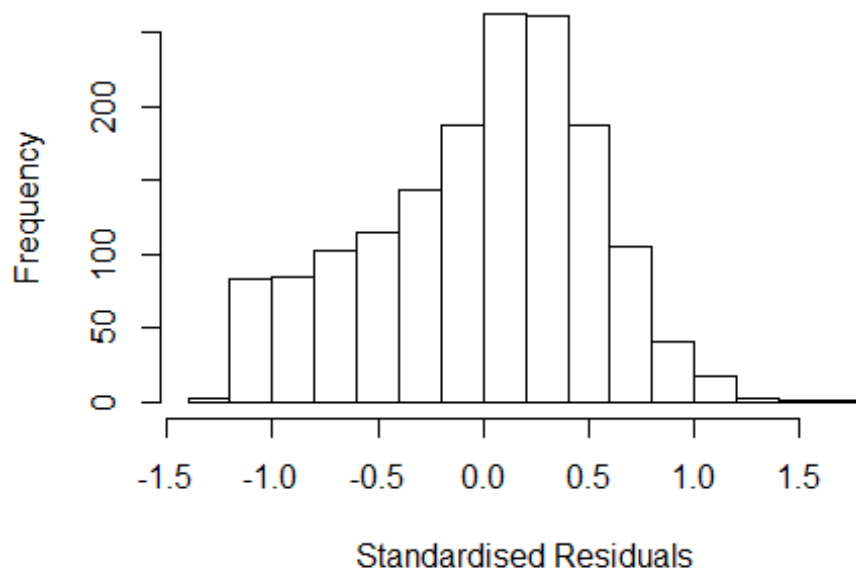
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3260 -0.3863  0.0605  0.3645  1.6095
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0527    0.0992   10.62  <2e-16 ***
## FirstAuthorFemale1  0.0580    0.0285    2.03  0.0422 *
## LastAuthorFemale1 -0.0831    0.0310   -2.68  0.0075 **
## Year1997          0.0332    0.1368    0.24  0.8085
## Year1998          0.2153    0.1288    1.67  0.0947 .
## Year1999          0.0225    0.1294    0.17  0.8619
## Year2000          0.0394    0.1137    0.35  0.7292
## Year2001          0.1411    0.1159    1.22  0.2236
## Year2002          0.0638    0.1123    0.57  0.5703
## Year2003         -0.0379    0.1140   -0.33  0.7396
## Year2004          0.0178    0.1180    0.15  0.8800
## Year2005          0.0355    0.1069    0.33  0.7398
```

```

## Year2006          -0.0453      0.1090   -0.42   0.6775
## Year2007          -0.0766      0.1149   -0.67   0.5049
## Year2008          -0.0690      0.1112   -0.62   0.5349
## Year2009          -0.0819      0.1148   -0.71   0.4756
## Year2010           0.0667      0.1074    0.62   0.5348
## Year2011           0.0399      0.1076    0.37   0.7110
## Year2012          -0.0426      0.1099   -0.39   0.6981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0221, Adjusted R-squared:  0.011
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1479 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.319  0.863   0.945   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      6.25e-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 1.057 1      1.028
## Year              1.057 16      1.002

```


Residuals from first author



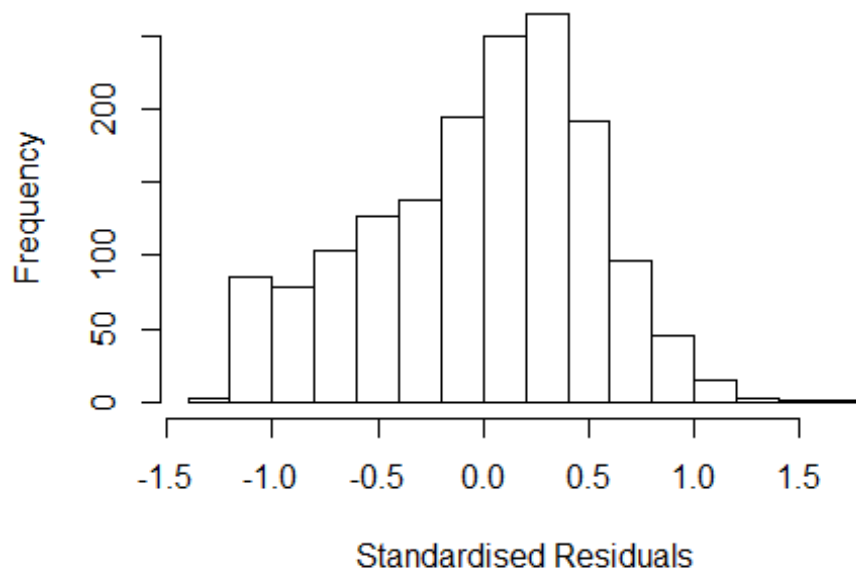
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.298 -0.385  0.066  0.357  1.628
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0391     0.0961   10.81  <2e-16 ***
## FirstAuthorFemale1  0.0419     0.0287    1.46   0.145
## Year1997          0.0393     0.1347    0.29   0.770
## Year1998          0.2171     0.1267    1.71   0.087 .
## Year1999          0.0195     0.1263    0.15   0.878
## Year2000          0.0430     0.1113    0.39   0.699
## Year2001          0.1337     0.1142    1.17   0.242
## Year2002          0.0559     0.1104    0.51   0.613
## Year2003         -0.0498     0.1120   -0.44   0.657
## Year2004          0.0156     0.1158    0.13   0.893
## Year2005          0.0263     0.1041    0.25   0.801
## Year2006         -0.0563     0.1066   -0.53   0.597
```

```

## Year2007          -0.0824      0.1125   -0.73    0.464
## Year2008          -0.0761      0.1090   -0.70    0.485
## Year2009          -0.0948      0.1125   -0.84    0.399
## Year2010           0.0580      0.1047    0.55    0.579
## Year2011           0.0351      0.1049    0.33    0.738
## Year2012          -0.0470      0.1072   -0.44    0.661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0174, Adjusted R-squared:  0.00685
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1480 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.860  0.946  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      6.25e-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 1.035 1          1.018
## Year              1.035 16          1.001

```

Residuals from last author



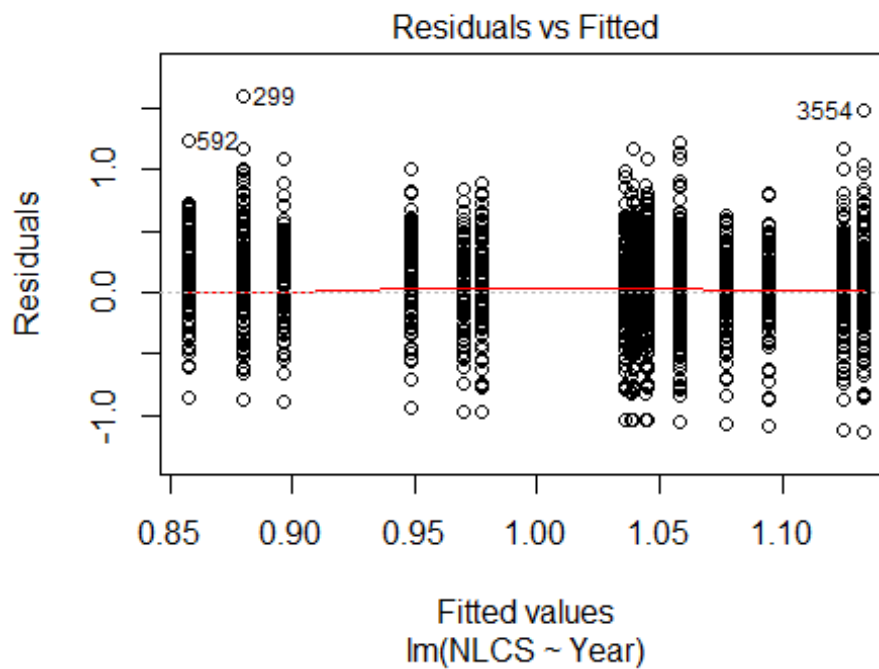
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2871 -0.3887 0.0578 0.3637 1.5827
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0683 0.0985 10.84 <2e-16 ***
## LastAuthorFemale1 -0.0704 0.0310 -2.27 0.023 *
## Year1997 0.0423 0.1375 0.31 0.758
## Year1998 0.2188 0.1290 1.70 0.090 .
## Year1999 0.0193 0.1289 0.15 0.881
## Year2000 0.0368 0.1133 0.32 0.745
## Year2001 0.1465 0.1158 1.27 0.206
## Year2002 0.0694 0.1128 0.62 0.538
## Year2003 -0.0324 0.1138 -0.28 0.776
## Year2004 0.0254 0.1177 0.22 0.829
## Year2005 0.0475 0.1067 0.45 0.656
## Year2006 -0.0414 0.1090 -0.38 0.704
```

```

## Year2007          -0.0654      0.1150   -0.57    0.569
## Year2008          -0.0579      0.1112   -0.52    0.602
## Year2009          -0.0712      0.1150   -0.62    0.536
## Year2010           0.0731      0.1074    0.68    0.496
## Year2011           0.0510      0.1076    0.47    0.636
## Year2012          -0.0312      0.1099   -0.28    0.777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0197, Adjusted R-squared:  0.00915
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1476 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.334  0.861  0.946  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      6.25e-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: 1600"
## [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
## 168 211 180 181 175 165 143 157 151 144 169 164 172 241 232
## 2011 2012
## 250 283
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 103 128 110 110 102 81 92 103 96 104 119 122 125 168 157
## 2011 2012

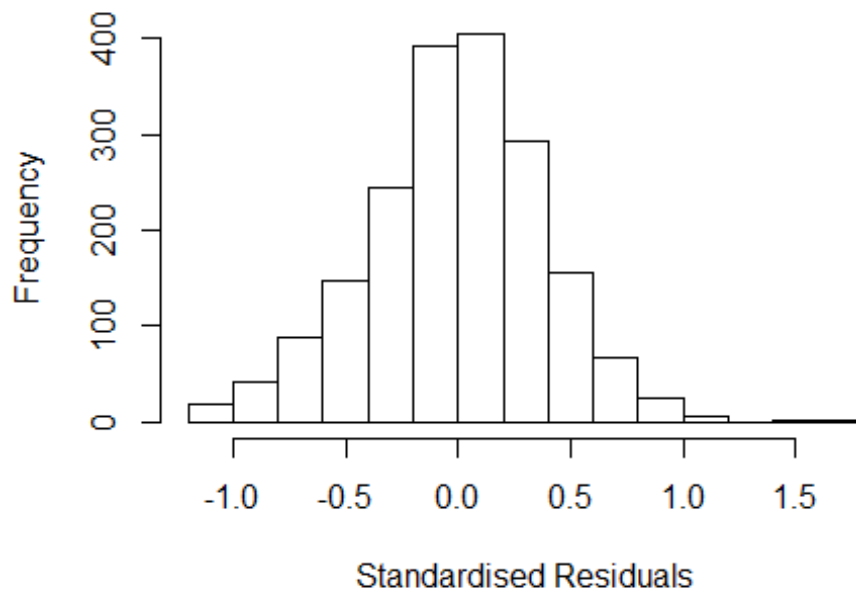
```

```
## 181 206
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 97 107 95 102 89 70 80 88 83 96 112 108 113 150 142
## 2011 2012
## 165 194
## [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 = 42, df = 1, p-value = 1e-10
```


Residuals from first and last author and team size



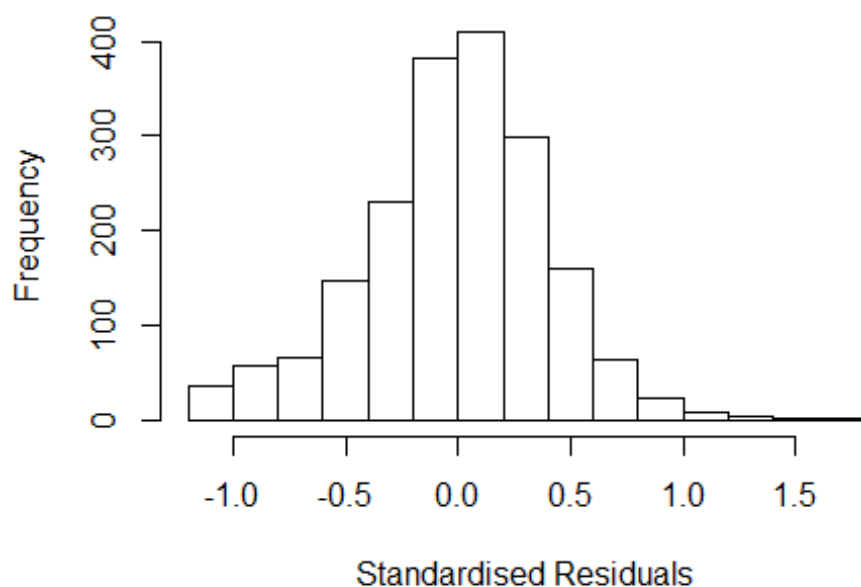
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19249 -0.24207  0.00301  0.24526  1.64586
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.71600    0.07075   10.12 < 2e-16 ***
## FirstAuthorFemale1 0.04704    0.01833    2.57 0.01034 *
## LastAuthorFemale1 -0.00649    0.02343   -0.28 0.78171
## UniqueAuthors2    0.21084    0.06008    3.51 0.00046 ***
## UniqueAuthors3    0.22083    0.05889    3.75 0.00018 ***
## UniqueAuthors4    0.28665    0.05958    4.81 1.6e-06 ***
## UniqueAuthors5    0.40125    0.05693    7.05 2.5e-12 ***
## Year1997         -0.10469    0.07197   -1.45 0.14596
## Year1998         -0.10324    0.07262   -1.42 0.15526
## Year1999         -0.08854    0.06162   -1.44 0.15091
```

```

## Year2000          0.07988      0.05962      1.34  0.18045
## Year2001         -0.04401      0.07165     -0.61  0.53913
## Year2002          0.04366      0.06008      0.73  0.46747
## Year2003          0.04782      0.05783      0.83  0.40835
## Year2004         -0.02014      0.06537     -0.31  0.75805
## Year2005          0.01374      0.05929      0.23  0.81676
## Year2006          0.08604      0.05860      1.47  0.14221
## Year2007          0.07782      0.05812      1.34  0.18075
## Year2008          0.00886      0.06174      0.14  0.88592
## Year2009          0.02308      0.05924      0.39  0.69684
## Year2010          0.01332      0.05908      0.23  0.82163
## Year2011          0.07525      0.05884      1.28  0.20112
## Year2012          0.10613      0.05518      1.92  0.05460 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.115, Adjusted R-squared:  0.104
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1717 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8570 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          5.29e-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 1.048 1          1.024
## LastAuthorFemale  1.055 1          1.027
## Year              1.067 16          1.002

```


Residuals from first and last author



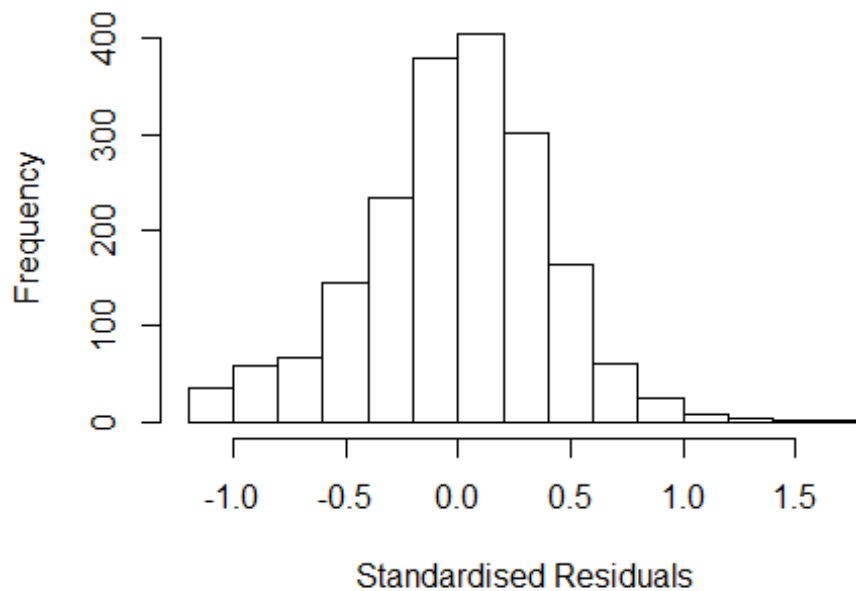
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.160 -0.248 0.012 0.247 1.636
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95904 0.05228 18.34 <2e-16 ***
## FirstAuthorFemale1 0.05900 0.01882 3.14 0.0017 **
## LastAuthorFemale1 -0.00935 0.02363 -0.40 0.6925
## Year1997 -0.11687 0.07401 -1.58 0.1145
## Year1998 -0.11167 0.07499 -1.49 0.1366
## Year1999 -0.06759 0.06401 -1.06 0.2912
## Year2000 0.13535 0.06238 2.17 0.0302 *
## Year2001 -0.01407 0.08054 -0.17 0.8613
## Year2002 0.07595 0.06272 1.21 0.2261
## Year2003 0.09203 0.06150 1.50 0.1347
## Year2004 0.01914 0.07069 0.27 0.7866
## Year2005 0.08634 0.06332 1.36 0.1729
```

```

## Year2006          0.14171      0.06131      2.31      0.0209 *
## Year2007          0.14356      0.06189      2.32      0.0205 *
## Year2008          0.06448      0.06540      0.99      0.3243
## Year2009          0.07510      0.06243      1.20      0.2292
## Year2010          0.07993      0.06333      1.26      0.2070
## Year2011          0.12710      0.06186      2.05      0.0400 *
## Year2012          0.15347      0.05821      2.64      0.0084 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0398
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~= 1. The remaining 1729 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.015  0.858   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      5.29e-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 1.029 1          1.015
## Year              1.029 16          1.001

```

Residuals from first author



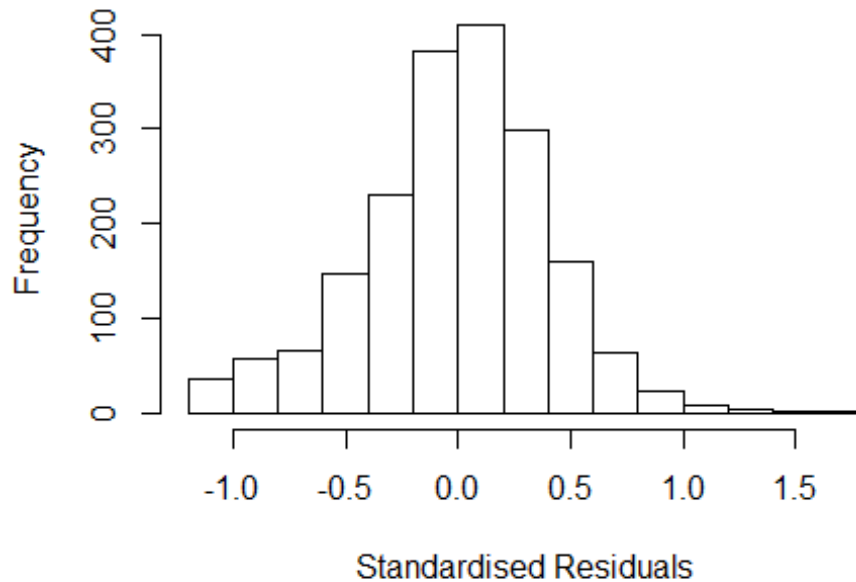
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1577 -0.2462 0.0113 0.2474 1.6381
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9580 0.0523 18.33 <2e-16 ***
## FirstAuthorFemale1 0.0581 0.0186 3.11 0.0019 **
## Year1997 -0.1181 0.0738 -1.60 0.1100
## Year1998 -0.1127 0.0750 -1.50 0.1330
## Year1999 -0.0673 0.0640 -1.05 0.2930
## Year2000 0.1353 0.0624 2.17 0.0304 *
## Year2001 -0.0141 0.0806 -0.17 0.8613
## Year2002 0.0765 0.0627 1.22 0.2222
## Year2003 0.0913 0.0615 1.49 0.1376
## Year2004 0.0195 0.0707 0.28 0.7831
## Year2005 0.0860 0.0633 1.36 0.1746
## Year2006 0.1416 0.0613 2.31 0.0210 *
```

```

## Year2007          0.1432      0.0619      2.32      0.0207 *
## Year2008          0.0637      0.0654      0.98      0.3295
## Year2009          0.0746      0.0624      1.20      0.2322
## Year2010          0.0794      0.0633      1.25      0.2098
## Year2011          0.1263      0.0618      2.04      0.0410 *
## Year2012          0.1528      0.0582      2.63      0.0087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0403
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1728 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0142 0.8580 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      5.29e-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 1.039 1      1.019
## Year              1.039 16      1.001

```

Residuals from last author



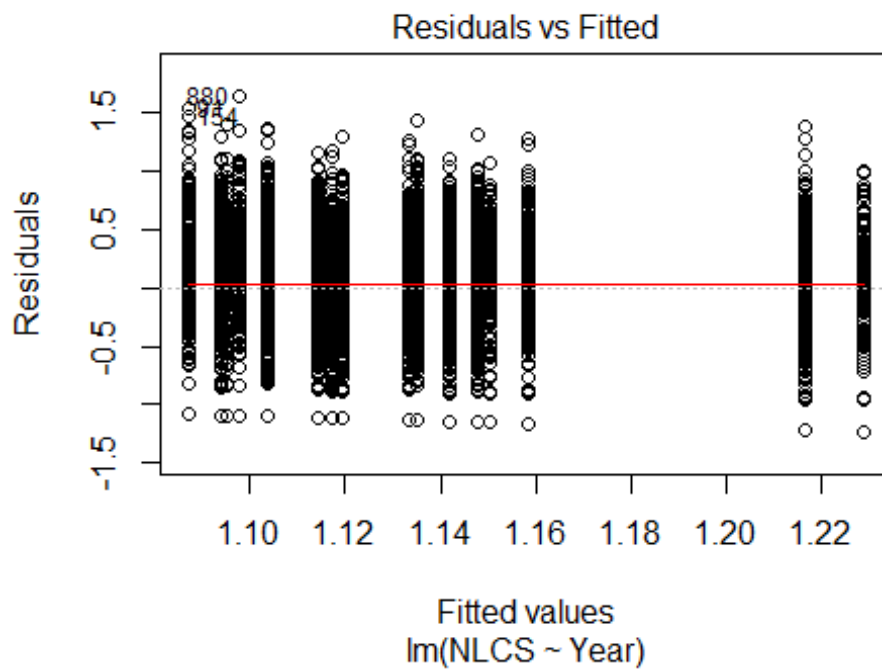
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1360 -0.2435 0.0145 0.2463 1.6196
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97505 0.05211 18.71 <2e-16 ***
## LastAuthorFemale1 -0.00055 0.02338 -0.02 0.9812
## Year1997 -0.11664 0.07411 -1.57 0.1157
## Year1998 -0.10582 0.07551 -1.40 0.1612
## Year1999 -0.05843 0.06432 -0.91 0.3637
## Year2000 0.14158 0.06238 2.27 0.0234 *
## Year2001 -0.01084 0.08140 -0.13 0.8941
## Year2002 0.08235 0.06269 1.31 0.1891
## Year2003 0.09082 0.06160 1.47 0.1406
## Year2004 0.02742 0.07048 0.39 0.6973
## Year2005 0.09274 0.06331 1.46 0.1431
## Year2006 0.14880 0.06163 2.41 0.0159 *
```

```

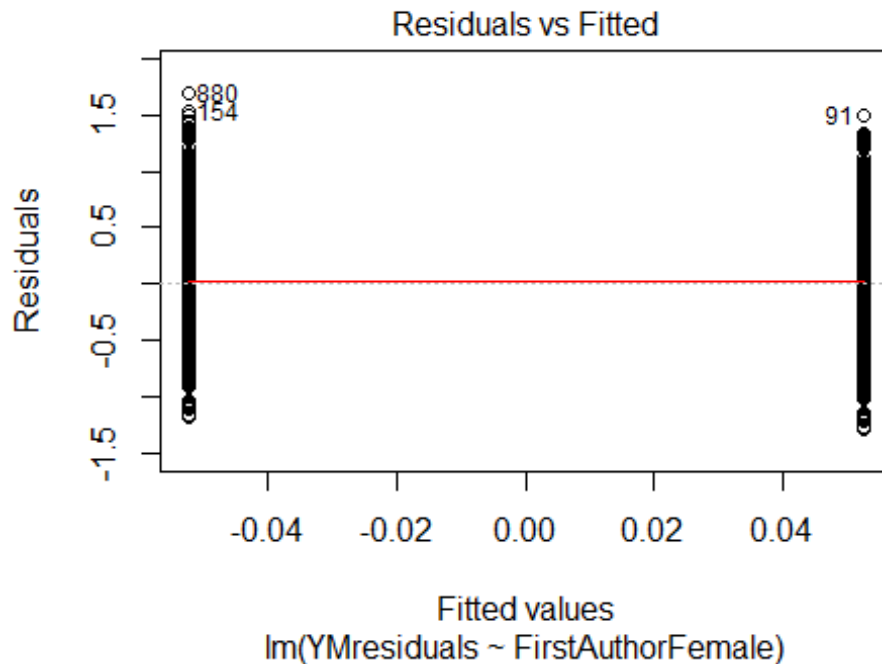
## Year2007      0.15446    0.06198    2.49    0.0128 *
## Year2008      0.07143    0.06549    1.09    0.2756
## Year2009      0.07883    0.06266    1.26    0.2085
## Year2010      0.08712    0.06335    1.38    0.1692
## Year2011      0.13363    0.06191    2.16    0.0310 *
## Year2012      0.16098    0.05818    2.77    0.0057 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.0444, Adjusted R-squared:  0.0357
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 167 weights are ~= 1. The remaining 1724 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0199 0.8550 0.9520 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      5.29e-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: 1891"
## [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
## 305 342 335 321 367 397 385 346 374 421 514 556 604 616 617
## 2011 2012
## 760 791
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 221 243 220 222 194 170 273 285 320 363 428 470 515 511 511
## 2011 2012

```

```
## 629 662
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 209 220 203 199 173 161 245 241 282 324 392 430 473 467 466
## 2011 2012
## 585 598
## [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
```

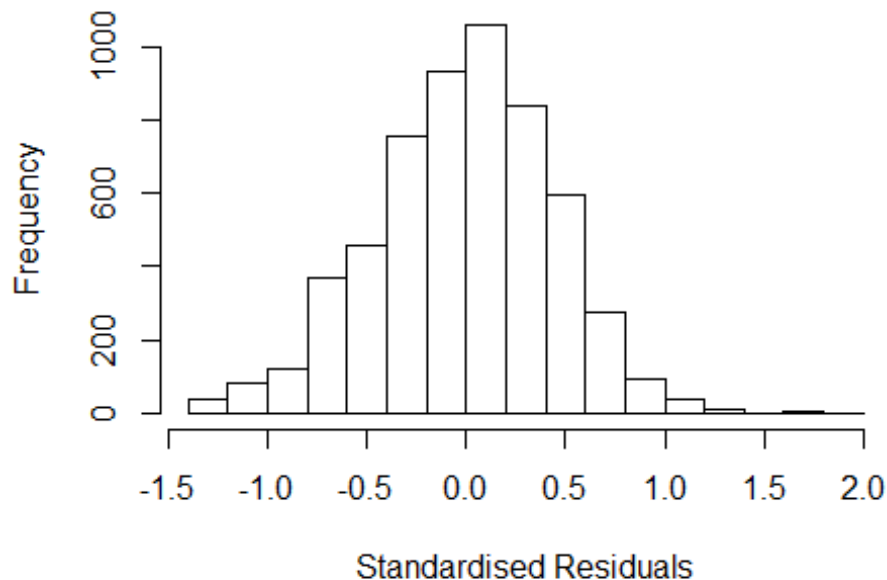


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 35, df = 1, p-value = 4e-09
```



```
## [1] "Female first author team size 2018 geometric mean: 4.17791483288063"
## [1] "Male first author team size 2018 geometric mean: 3.57595608146989"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 62000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.78918726768329"
## [1] "Male last author team size 2018 geometric mean: 4.0651438515686"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 53000, 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 1.056 1      1.028
## LastAuthorFemale  1.042 1      1.021
## UniqueAuthors    1.120 4      1.014
## Year              1.128 16     1.004
```


Residuals from first and last author and team size



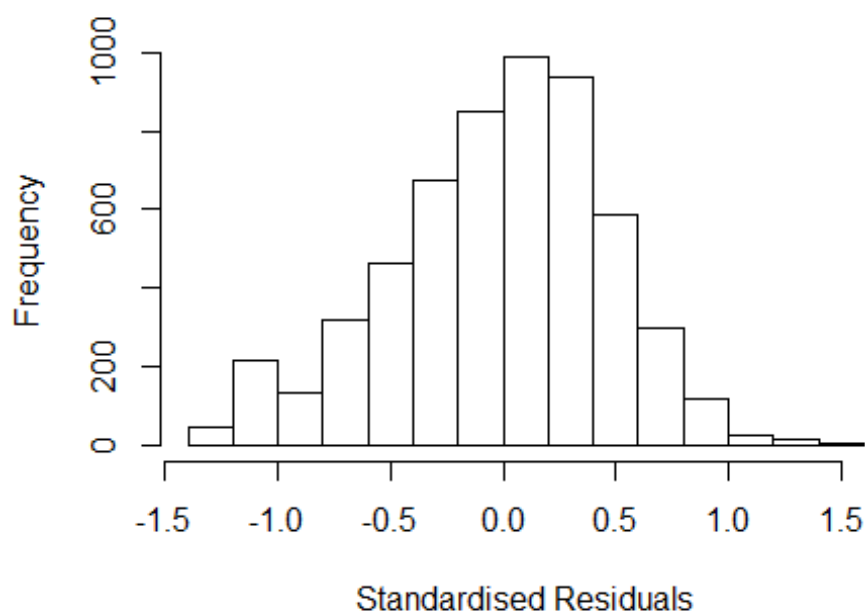
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3728 -0.3027 0.0167 0.2926 1.8124
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.77321 0.04446 17.39 <2e-16 ***
## FirstAuthorFemale1 0.06158 0.01240 4.97 7e-07 ***
## LastAuthorFemale1 0.02627 0.01249 2.10 0.0355 *
## UniqueAuthors2 0.39577 0.02329 16.99 <2e-16 ***
## UniqueAuthors3 0.47101 0.02315 20.34 <2e-16 ***
## UniqueAuthors4 0.51938 0.02465 21.07 <2e-16 ***
## UniqueAuthors5 0.57066 0.02250 25.36 <2e-16 ***
## Year1997 -0.07120 0.05401 -1.32 0.1874
## Year1998 -0.04592 0.05744 -0.80 0.4241
## Year1999 -0.10160 0.05493 -1.85 0.0644 .
```

```

## Year2000      -0.00704    0.05601   -0.13    0.8999
## Year2001      0.04070    0.05566    0.73    0.4646
## Year2002     -0.04521    0.05226   -0.87    0.3869
## Year2003     -0.09360    0.05145   -1.82    0.0689 .
## Year2004     -0.10297    0.04982   -2.07    0.0388 *
## Year2005     -0.10130    0.04889   -2.07    0.0383 *
## Year2006     -0.10330    0.04828   -2.14    0.0324 *
## Year2007     -0.10078    0.04740   -2.13    0.0335 *
## Year2008     -0.09372    0.04693   -2.00    0.0459 *
## Year2009     -0.07980    0.04707   -1.70    0.0901 .
## Year2010     -0.04088    0.04729   -0.86    0.3874
## Year2011     -0.10785    0.04660   -2.31    0.0207 *
## Year2012     -0.14644    0.04692   -3.12    0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.167, Adjusted R-squared:  0.164
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 465 weights are ~= 1. The remaining 5203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0486 0.8710 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      1.76e-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 1.029 1 1.015
## LastAuthorFemale 1.018 1 1.009
## Year 1.044 16 1.001

```

Residuals from first and last author



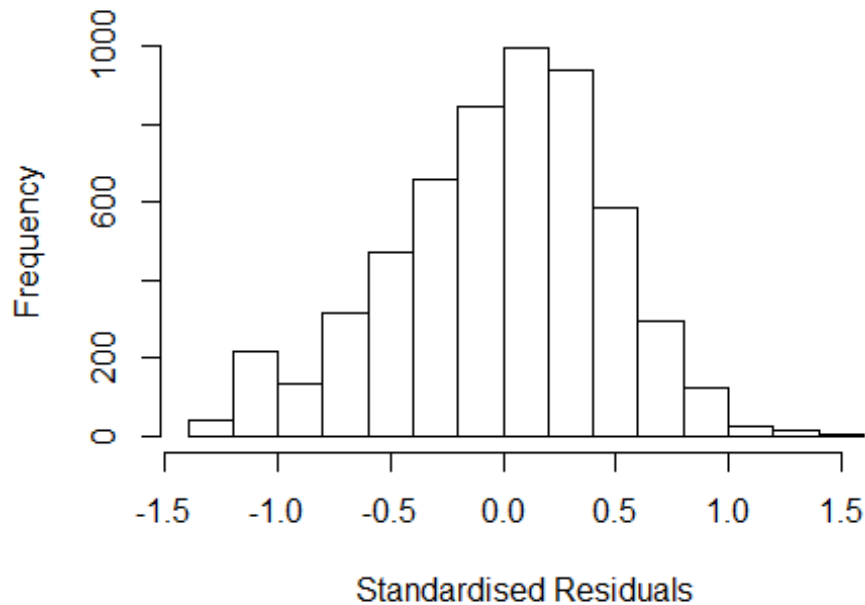
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3438 -0.3178  0.0296  0.3167  1.5159
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04311    0.04880   21.37 < 2e-16 ***
## FirstAuthorFemale1 0.10368    0.01321    7.85 5.1e-15 ***
## LastAuthorFemale1 0.03054    0.01323    2.31 0.0210 *
## Year1997        -0.01012    0.06210   -0.16 0.8705
## Year1998         0.01068    0.06290    0.17 0.8652
## Year1999        -0.00215    0.06145   -0.03 0.9721
## Year2000         0.08214    0.06107    1.34 0.1787
## Year2001         0.16644    0.06217    2.68 0.0075 **
## Year2002         0.07515    0.05770    1.30 0.1928
## Year2003         0.01225    0.05778    0.21 0.8321
## Year2004         0.00652    0.05607    0.12 0.9074
## Year2005         0.03269    0.05505    0.59 0.5526
```

```

## Year2006          0.04673      0.05428      0.86      0.3893
## Year2007          0.02970      0.05334      0.56      0.5776
## Year2008          0.03856      0.05301      0.73      0.4671
## Year2009          0.04666      0.05336      0.87      0.3819
## Year2010          0.13167      0.05297      2.49      0.0130 *
## Year2011          0.04252      0.05275      0.81      0.4203
## Year2012          0.00436      0.05293      0.08      0.9343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.462
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.0192
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 478 weights are ~= 1. The remaining 5190 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.260  0.866  0.949  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      1.76e-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 1.027 1      1.013
## Year              1.027 16      1.001

```

Residuals from first author



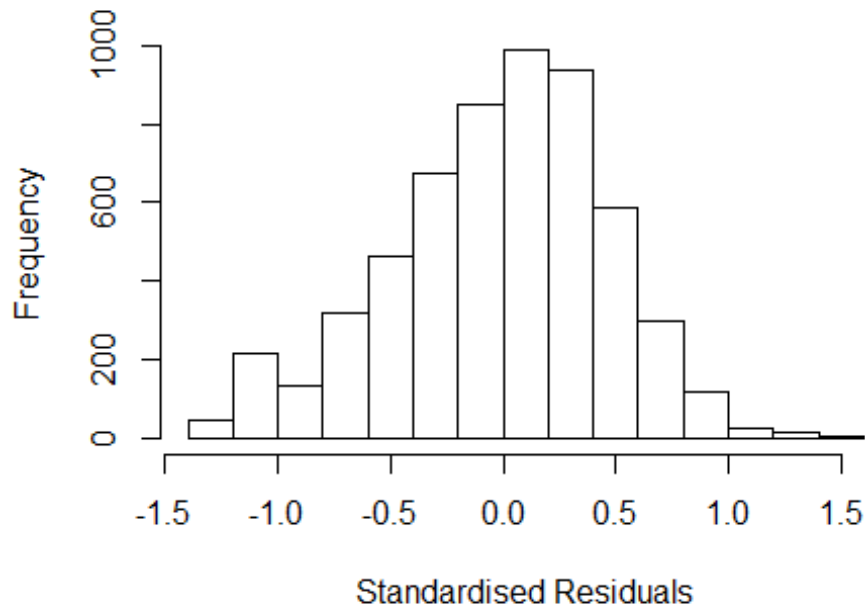
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3312 -0.3185 0.0288 0.3178 1.5058
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05321 0.04880 21.58 < 2e-16 ***
## FirstAuthorFemale1 0.10813 0.01328 8.14 4.7e-16 ***
## Year1997 -0.00978 0.06237 -0.16 0.8754
## Year1998 0.00627 0.06302 0.10 0.9208
## Year1999 -0.00234 0.06156 -0.04 0.9697
## Year2000 0.08365 0.06122 1.37 0.1719
## Year2001 0.16981 0.06221 2.73 0.0064 **
## Year2002 0.07519 0.05789 1.30 0.1941
## Year2003 0.01444 0.05789 0.25 0.8030
## Year2004 0.00330 0.05623 0.06 0.9532
## Year2005 0.03313 0.05518 0.60 0.5483
## Year2006 0.04673 0.05444 0.86 0.3907
```

```

## Year2007          0.02939    0.05354    0.55    0.5830
## Year2008          0.03848    0.05319    0.72    0.4695
## Year2009          0.04647    0.05354    0.87    0.3855
## Year2010          0.13304    0.05311    2.51    0.0123 *
## Year2011          0.04268    0.05292    0.81    0.4200
## Year2012          0.00326    0.05309    0.06    0.9510
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.462
## Multiple R-squared:  0.0213, Adjusted R-squared:  0.0183
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 487 weights are ~= 1. The remaining 5181 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.267  0.867  0.949  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      1.76e-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 1.015 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
## -1.3015 -0.3255  0.0343  0.3186  1.5594
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06956    0.04820   22.19  < 2e-16 ***
## LastAuthorFemale1 0.04605    0.01331    3.46  0.00054 ***
## Year1997      -0.00259    0.06176   -0.04  0.96656
## Year1998       0.01758    0.06338    0.28  0.78152
## Year1999       0.00716    0.06142    0.12  0.90721
## Year2000       0.09819    0.06058    1.62  0.10513
## Year2001       0.18592    0.06166    3.02  0.00258 **
## Year2002       0.09079    0.05732    1.58  0.11323
## Year2003       0.02017    0.05744    0.35  0.72553
## Year2004       0.02746    0.05548    0.49  0.62073
## Year2005       0.05436    0.05432    1.00  0.31701
## Year2006       0.06856    0.05378    1.27  0.20241
```

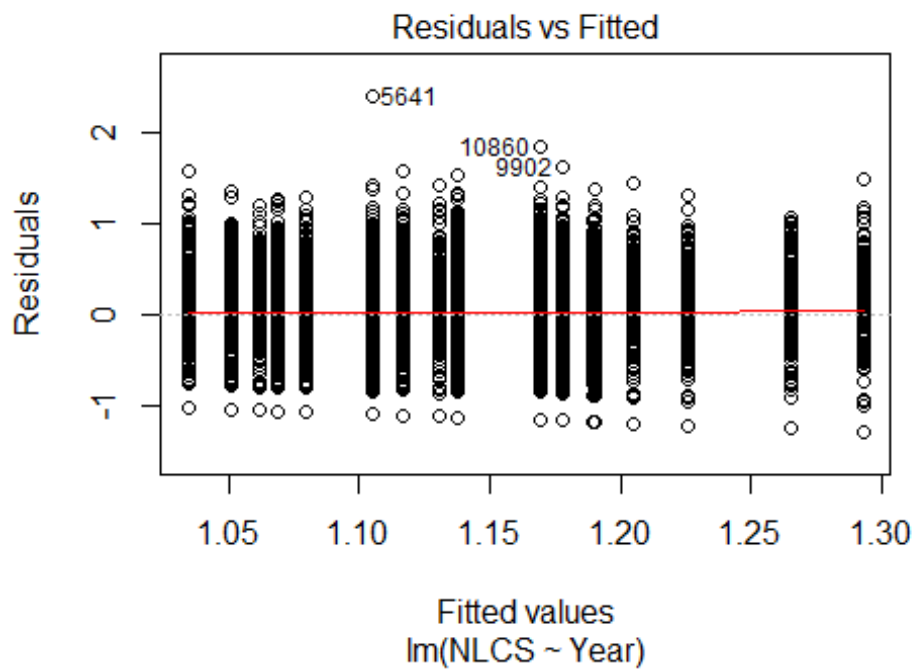
```

## Year2007      0.05095      0.05272      0.97  0.33391
## Year2008      0.05953      0.05247      1.13  0.25657
## Year2009      0.07103      0.05270      1.35  0.17778
## Year2010      0.15457      0.05238      2.95  0.00318 **
## Year2011      0.07216      0.05193      1.39  0.16466
## Year2012      0.03282      0.05230      0.63  0.53028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.464
## Multiple R-squared:  0.0111, Adjusted R-squared:  0.00812
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 440 weights are ~= 1. The remaining 5228 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.236  0.867  0.950  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      1.76e-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: 5668"
## [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
## 327 348 369 316 411 448 486 434 414 509 661 656 744 804 811
## 2011 2012
## 955 964
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 193 207 219 163 200 205 345 314 304 416 510 528 598 627 634
## 2011 2012

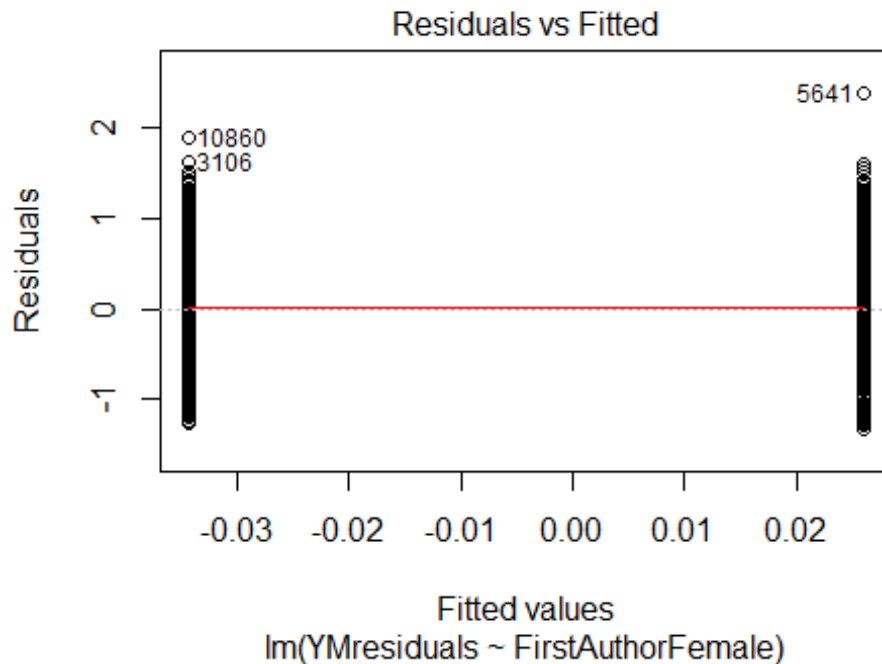
```



```
## 719 714
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 182 188 199 149 178 177 316 284 269 377 459 475 526 568 575
## 2011 2012
## 651 647
## [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
```

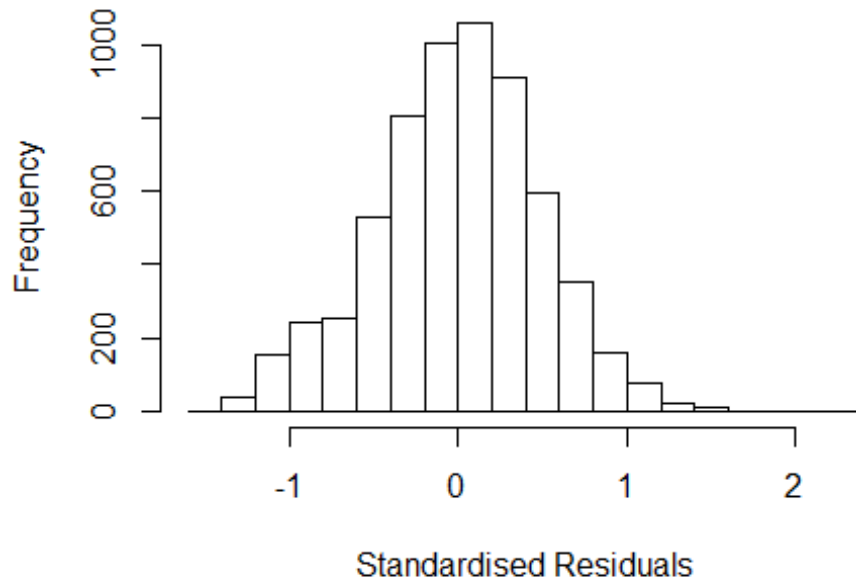


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.09821615092959"
## [1] "Male first author team size 2018 geometric mean: 3.77476867350282"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 95000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.93519905791968"
## [1] "Male last author team size 2018 geometric mean: 4.06043207780199"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 99000, 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 1.094 1      1.046
## LastAuthorFemale  1.093 1      1.045
## UniqueAuthors    1.111 4      1.013
## Year             1.124 16      1.004
```

Residuals from first and last author and team size



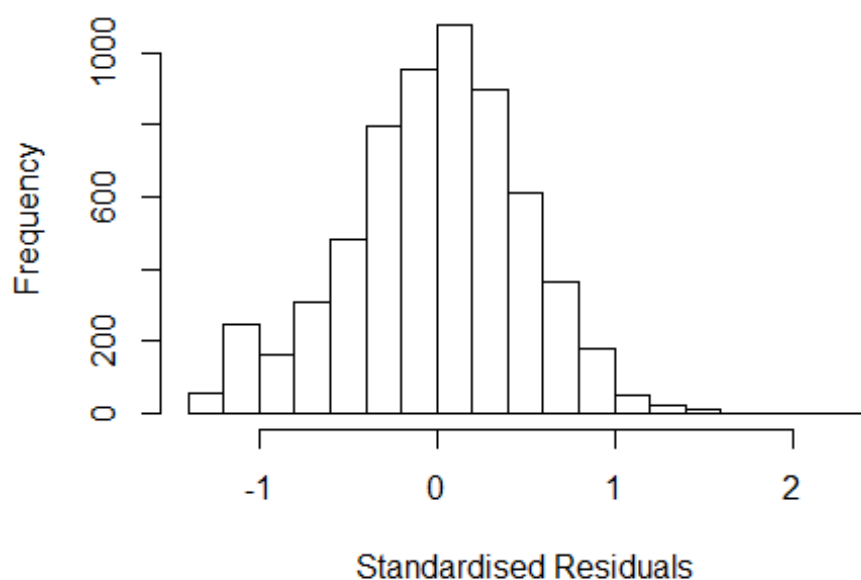
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4107 -0.3132 0.0137 0.3132 2.3603
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09379 0.04393 24.90 < 2e-16 ***
## FirstAuthorFemale1 0.03571 0.01332 2.68 0.00738 **
## LastAuthorFemale1 0.00727 0.01305 0.56 0.57778
## UniqueAuthors2 0.18326 0.02429 7.54 5.2e-14 ***
## UniqueAuthors3 0.27297 0.02346 11.63 < 2e-16 ***
## UniqueAuthors4 0.31745 0.02401 13.22 < 2e-16 ***
## UniqueAuthors5 0.36303 0.02244 16.18 < 2e-16 ***
## Year1997 -0.04615 0.05599 -0.82 0.40984
## Year1998 -0.08833 0.05422 -1.63 0.10334
## Year1999 -0.07831 0.05670 -1.38 0.16725
```

```

## Year2000      -0.11154      0.05388      -2.07      0.03850 *
## Year2001      -0.21908      0.05728      -3.82      0.00013 ***
## Year2002      -0.28347      0.05094      -5.56      2.7e-08 ***
## Year2003      -0.27549      0.05085      -5.42      6.3e-08 ***
## Year2004      -0.28458      0.05081      -5.60      2.2e-08 ***
## Year2005      -0.27015      0.04806      -5.62      2.0e-08 ***
## Year2006      -0.27620      0.04698      -5.88      4.4e-09 ***
## Year2007      -0.25578      0.04630      -5.52      3.4e-08 ***
## Year2008      -0.23591      0.04642      -5.08      3.8e-07 ***
## Year2009      -0.22922      0.04503      -5.09      3.7e-07 ***
## Year2010      -0.19386      0.04491      -4.32      1.6e-05 ***
## Year2011      -0.19720      0.04474      -4.41      1.1e-05 ***
## Year2012      -0.21747      0.04488      -4.85      1.3e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0799, Adjusted R-squared:  0.0766
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2809 is an outlier with |weight| = 0 ( < 1.6e-05);
## 516 weights are ~= 1. The remaining 5703 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0019 0.8670 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      1.61e-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 1.073 1      1.036
## LastAuthorFemale 1.065 1      1.032
## Year      1.044 16      1.001

```

Residuals from first and last author



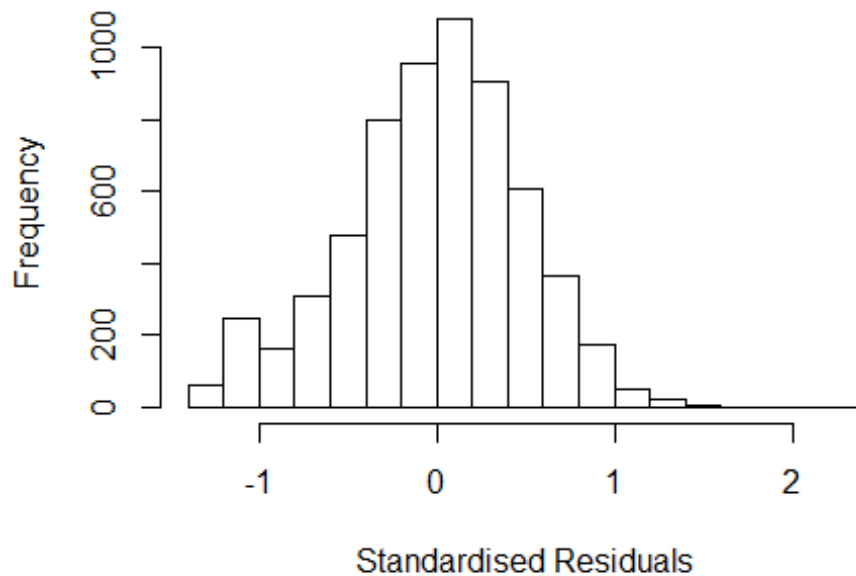
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3393 -0.3220 0.0174 0.3235 2.3641
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28127 0.03945 32.48 < 2e-16 ***
## FirstAuthorFemale1 0.06604 0.01366 4.83 1.4e-06 ***
## LastAuthorFemale1 -0.00806 0.01330 -0.61 0.5445
## Year1997 -0.02873 0.05539 -0.52 0.6040
## Year1998 -0.10332 0.05370 -1.92 0.0544 .
## Year1999 -0.07268 0.05602 -1.30 0.1946
## Year2000 -0.07212 0.05283 -1.37 0.1722
## Year2001 -0.16925 0.05567 -3.04 0.0024 **
## Year2002 -0.27443 0.05177 -5.30 1.2e-07 ***
## Year2003 -0.24389 0.05080 -4.80 1.6e-06 ***
## Year2004 -0.25010 0.05091 -4.91 9.2e-07 ***
## Year2005 -0.23093 0.04685 -4.93 8.5e-07 ***
```

```

## Year2006          -0.23430      0.04604      -5.09   3.7e-07 ***
## Year2007          -0.20437      0.04519      -4.52   6.2e-06 ***
## Year2008          -0.18404      0.04514      -4.08   4.6e-05 ***
## Year2009          -0.17840      0.04363      -4.09   4.4e-05 ***
## Year2010          -0.12570      0.04340      -2.90   0.0038 **
## Year2011          -0.12213      0.04322      -2.83   0.0047 **
## Year2012          -0.14072      0.04333      -3.25   0.0012 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.0183
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2809 is an outlier with |weight| = 0 ( < 1.6e-05);
## 516 weights are ~= 1. The remaining 5703 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0857 0.8640 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          1.61e-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 1.028 1          1.014
## Year              1.028 16          1.001

```

Residuals from first author



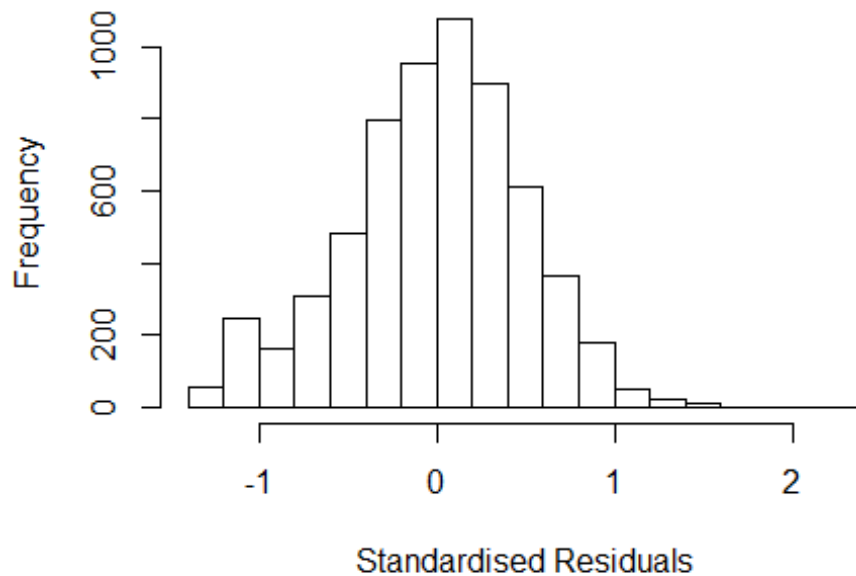
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3433 -0.3206 0.0178 0.3255 2.3688
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2794 0.0392 32.60 < 2e-16 ***
## FirstAuthorFemale1 0.0639 0.0134 4.77 1.9e-06 ***
## Year1997 -0.0288 0.0554 -0.52 0.6034
## Year1998 -0.1037 0.0536 -1.93 0.0533 .
## Year1999 -0.0730 0.0560 -1.30 0.1924
## Year2000 -0.0723 0.0528 -1.37 0.1707
## Year2001 -0.1700 0.0556 -3.06 0.0022 **
## Year2002 -0.2752 0.0517 -5.32 1.1e-07 ***
## Year2003 -0.2443 0.0508 -4.81 1.5e-06 ***
## Year2004 -0.2506 0.0509 -4.92 8.7e-07 ***
## Year2005 -0.2316 0.0468 -4.95 7.7e-07 ***
## Year2006 -0.2343 0.0460 -5.09 3.6e-07 ***
```

```

## Year2007          -0.2051      0.0451   -4.54  5.6e-06 ***
## Year2008          -0.1844      0.0451   -4.09  4.4e-05 ***
## Year2009          -0.1790      0.0436   -4.11  4.1e-05 ***
## Year2010          -0.1263      0.0434   -2.91   0.0036 **
## Year2011          -0.1228      0.0432   -2.85   0.0044 **
## Year2012          -0.1413      0.0433   -3.27   0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0211, Adjusted R-squared:  0.0184
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2809 is an outlier with |weight| = 0 ( < 1.6e-05);
## 511 weights are ~= 1. The remaining 5708 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0845 0.8640 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      1.61e-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 1.021 1          1.011
## Year          1.021 16          1.001

```


Residuals from last author



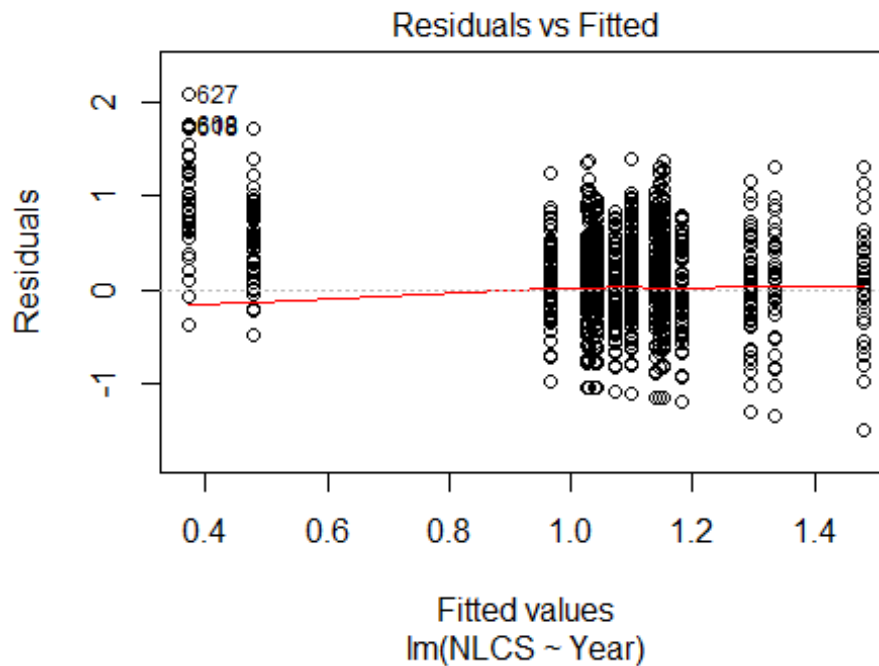
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3180 -0.3203 0.0197 0.3269 2.3998
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30894 0.03915 33.44 < 2e-16 ***
## LastAuthorFemale1 0.00911 0.01306 0.70 0.4856
## Year1997 -0.03115 0.05541 -0.56 0.5740
## Year1998 -0.10532 0.05385 -1.96 0.0506 .
## Year1999 -0.08139 0.05592 -1.46 0.1456
## Year2000 -0.07870 0.05295 -1.49 0.1372
## Year2001 -0.16974 0.05587 -3.04 0.0024 **
## Year2002 -0.27399 0.05189 -5.28 1.3e-07 ***
## Year2003 -0.24490 0.05079 -4.82 1.5e-06 ***
## Year2004 -0.24892 0.05099 -4.88 1.1e-06 ***
## Year2005 -0.23132 0.04688 -4.93 8.3e-07 ***
## Year2006 -0.23211 0.04607 -5.04 4.8e-07 ***
```

```

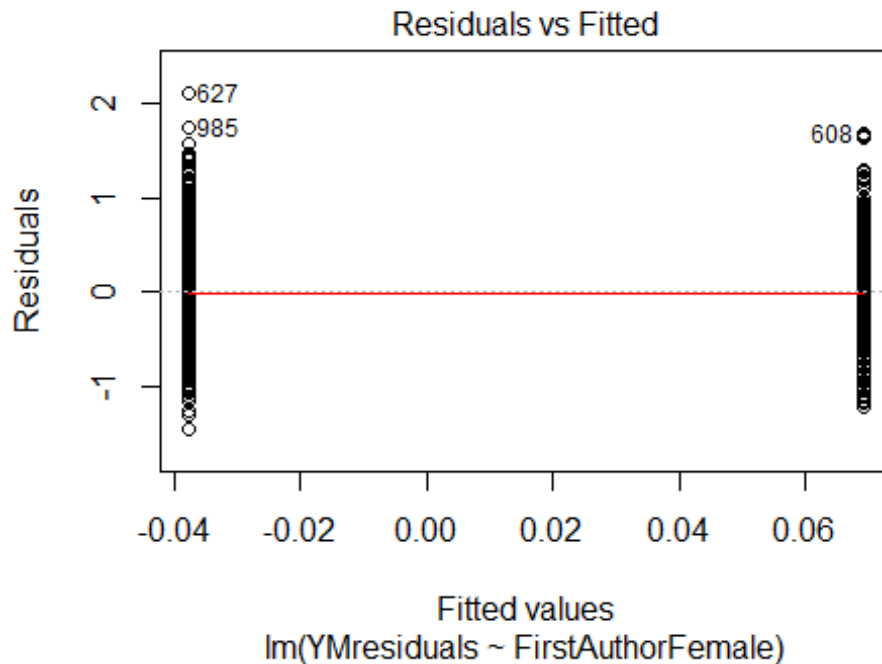
## Year2007          -0.20176      0.04528      -4.46      8.5e-06 ***
## Year2008          -0.18135      0.04519      -4.01      6.1e-05 ***
## Year2009          -0.17344      0.04375      -3.96      7.4e-05 ***
## Year2010          -0.11867      0.04352      -2.73      0.0064 **
## Year2011          -0.11484      0.04329      -2.65      0.0080 **
## Year2012          -0.13306      0.04337      -3.07      0.0022 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.0146
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2809 is an outlier with |weight| = 0 ( < 1.6e-05);
## 531 weights are ~= 1. The remaining 5688 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.863  0.950  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      1.61e-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: 6220"
## [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
##   90   90  103  272  113  256  116  105  126  135  128  137  185  200  224
## 2011 2012
##  232  213
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   51   42  140   61  114   70   70   78   97   89  106  141  150  161

```

```
## 2011 2012
## 185 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 32 45 37 127 51 103 61 60 70 81 78 94 132 136 144
## 2011 2012
## 171 146
## [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
```

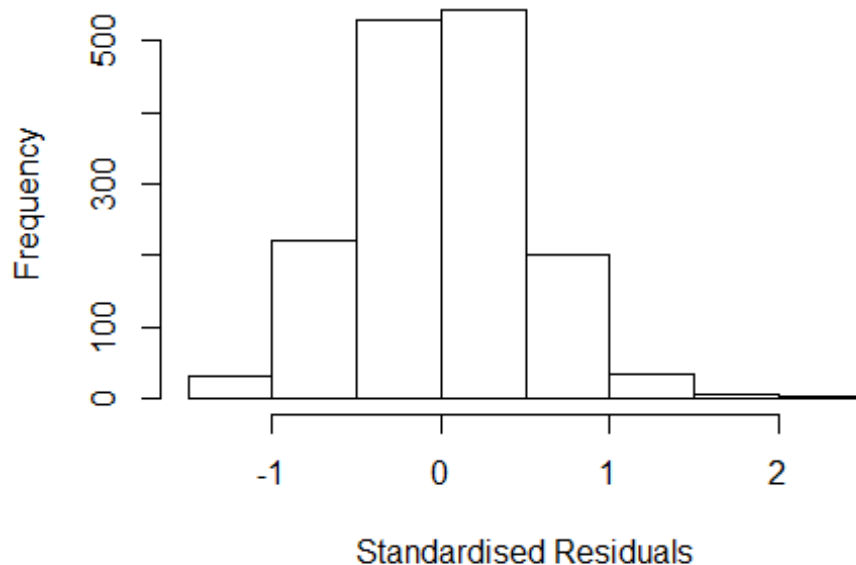


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.5, df = 1, p-value = 0.006
```



```
## [1] "Female first author team size 2018 geometric mean: 5.85708950655208"
## [1] "Male first author team size 2018 geometric mean: 4.93044189638297"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.77414020708934"
## [1] "Male last author team size 2018 geometric mean: 5.11176143406775"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, 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 1.150 1          1.072
## LastAuthorFemale  1.138 1          1.067
## UniqueAuthors    1.511 4          1.053
## Year              1.539 16         1.014
```

Residuals from first and last author and team size



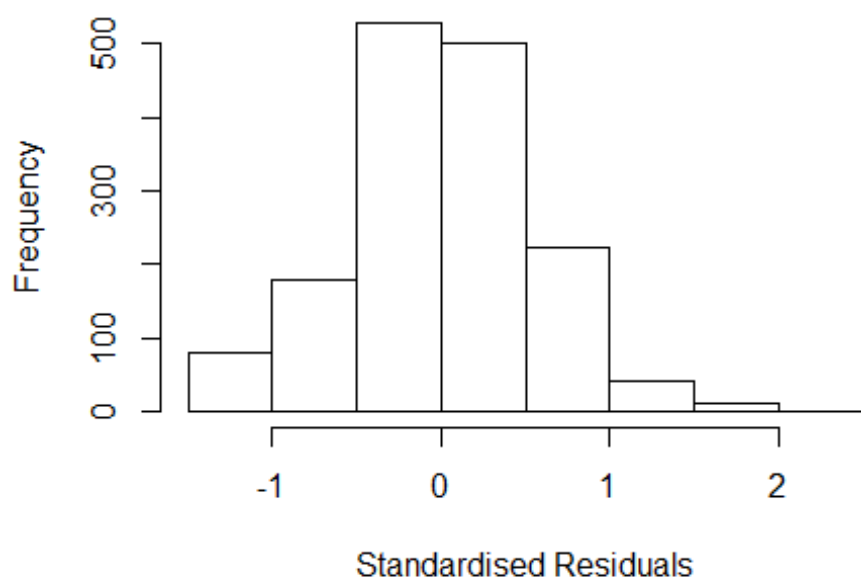
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.45169 -0.33584 0.00565 0.34336 2.41508
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0008 0.1432 6.99 4.1e-12 ***
## FirstAuthorFemale1 0.0922 0.0282 3.27 0.00112 **
## LastAuthorFemale1 0.0726 0.0324 2.25 0.02491 *
## UniqueAuthors2 0.3449 0.0677 5.09 3.9e-07 ***
## UniqueAuthors3 0.3461 0.0615 5.63 2.1e-08 ***
## UniqueAuthors4 0.4509 0.0599 7.53 8.4e-14 ***
## UniqueAuthors5 0.6830 0.0572 11.94 < 2e-16 ***
## Year1997 -0.1368 0.1524 -0.90 0.36957
## Year1998 -0.1447 0.1672 -0.87 0.38717
## Year1999 -1.3190 0.1414 -9.33 < 2e-16 ***
```

```

## Year2000          -0.3798      0.1572    -2.42   0.01582 *
## Year2001          -1.0559      0.1488    -7.10   1.9e-12 ***
## Year2002          -0.4514      0.1581    -2.86   0.00435 **
## Year2003          -0.4186      0.1472    -2.84   0.00451 **
## Year2004          -0.5247      0.1478    -3.55   0.00040 ***
## Year2005          -0.4558      0.1415    -3.22   0.00131 **
## Year2006          -0.5393      0.1438    -3.75   0.00018 ***
## Year2007          -0.4670      0.1418    -3.29   0.00101 **
## Year2008          -0.4894      0.1388    -3.53   0.00043 ***
## Year2009          -0.4095      0.1391    -2.94   0.00330 **
## Year2010          -0.4934      0.1373    -3.59   0.00034 ***
## Year2011          -0.3929      0.1375    -2.86   0.00433 **
## Year2012          -0.4355      0.1396    -3.12   0.00184 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.365, Adjusted R-squared:  0.356
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 233 is an outlier with |weight| = 0 ( < 6.4e-05);
## 128 weights are ~= 1. The remaining 1439 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0259 0.8550 0.9430 0.8930 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          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 1.085 1          1.042
## LastAuthorFemale 1.082 1          1.040
## Year 1.089 16          1.003

```

Residuals from first and last author



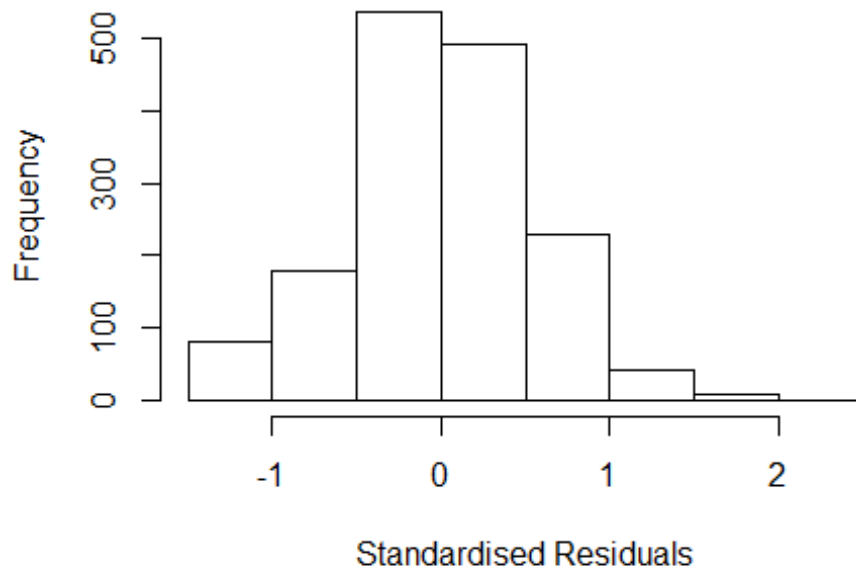
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46260 -0.31381 -0.00543 0.35774 2.33112
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4626 0.1464 9.99 < 2e-16 ***
## FirstAuthorFemale1 0.1226 0.0289 4.24 2.4e-05 ***
## LastAuthorFemale1 0.0430 0.0334 1.29 0.1979
## Year1997 -0.1639 0.1651 -0.99 0.3211
## Year1998 -0.1608 0.1872 -0.86 0.3905
## Year1999 -1.3507 0.1514 -8.92 < 2e-16 ***
## Year2000 -0.3369 0.1672 -2.02 0.0440 *
## Year2001 -1.1532 0.1615 -7.14 1.4e-12 ***
## Year2002 -0.3833 0.1740 -2.20 0.0278 *
## Year2003 -0.4334 0.1628 -2.66 0.0078 **
## Year2004 -0.5243 0.1604 -3.27 0.0011 **
## Year2005 -0.4508 0.1569 -2.87 0.0041 **
```

```

## Year2006          -0.4816      0.1570   -3.07   0.0022 **
## Year2007          -0.4502      0.1542   -2.92   0.0036 **
## Year2008          -0.4704      0.1544   -3.05   0.0024 **
## Year2009          -0.3486      0.1538   -2.27   0.0235 *
## Year2010          -0.4161      0.1504   -2.77   0.0057 **
## Year2011          -0.3321      0.1515   -2.19   0.0285 *
## Year2012          -0.3830      0.1531   -2.50   0.0124 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.277, Adjusted R-squared:  0.268
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 233 is an outlier with |weight| = 0 ( < 6.4e-05);
## 110 weights are ~= 1. The remaining 1457 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0922 0.8470 0.9460 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      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 1.045 1          1.022
## Year              1.045 16          1.001

```


Residuals from first author



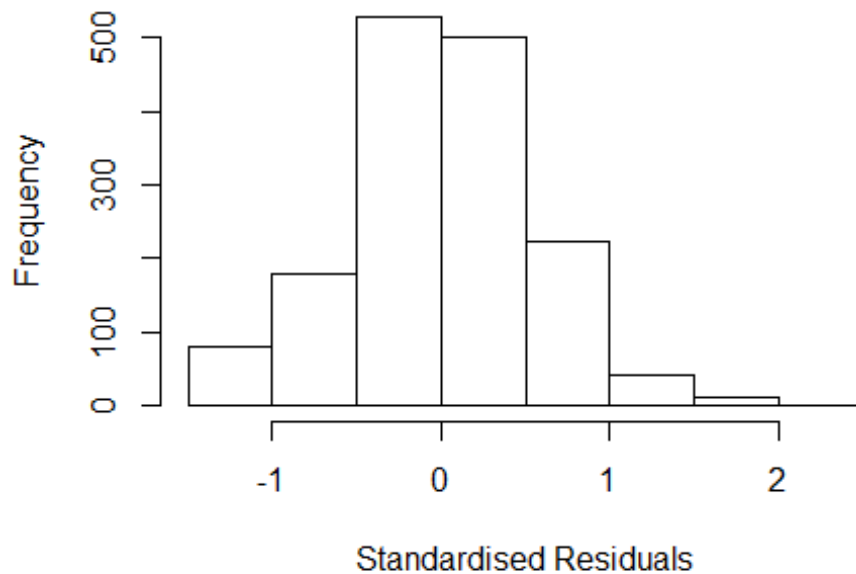
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46743 -0.31510 -0.00961 0.35527 2.32558
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4674 0.1459 10.06 < 2e-16 ***
## FirstAuthorFemale1 0.1311 0.0284 4.62 4.2e-06 ***
## Year1997 -0.1614 0.1645 -0.98 0.3266
## Year1998 -0.1644 0.1865 -0.88 0.3781
## Year1999 -1.3500 0.1511 -8.93 < 2e-16 ***
## Year2000 -0.3338 0.1669 -2.00 0.0456 *
## Year2001 -1.1523 0.1608 -7.17 1.2e-12 ***
## Year2002 -0.3892 0.1733 -2.25 0.0249 *
## Year2003 -0.4366 0.1624 -2.69 0.0072 **
## Year2004 -0.5248 0.1604 -3.27 0.0011 **
## Year2005 -0.4514 0.1566 -2.88 0.0040 **
## Year2006 -0.4812 0.1569 -3.07 0.0022 **
```

```

## Year2007          -0.4461      0.1538   -2.90   0.0038 **
## Year2008          -0.4682      0.1540   -3.04   0.0024 **
## Year2009          -0.3456      0.1534   -2.25   0.0244 *
## Year2010          -0.4167      0.1500   -2.78   0.0055 **
## Year2011          -0.3308      0.1511   -2.19   0.0287 *
## Year2012          -0.3802      0.1527   -2.49   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.276, Adjusted R-squared:  0.268
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 233 is an outlier with |weight| = 0 ( < 6.4e-05);
## 115 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0977 0.8430 0.9470 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          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 1.044 1          1.022
## Year          1.044 16          1.001

```

Residuals from last author



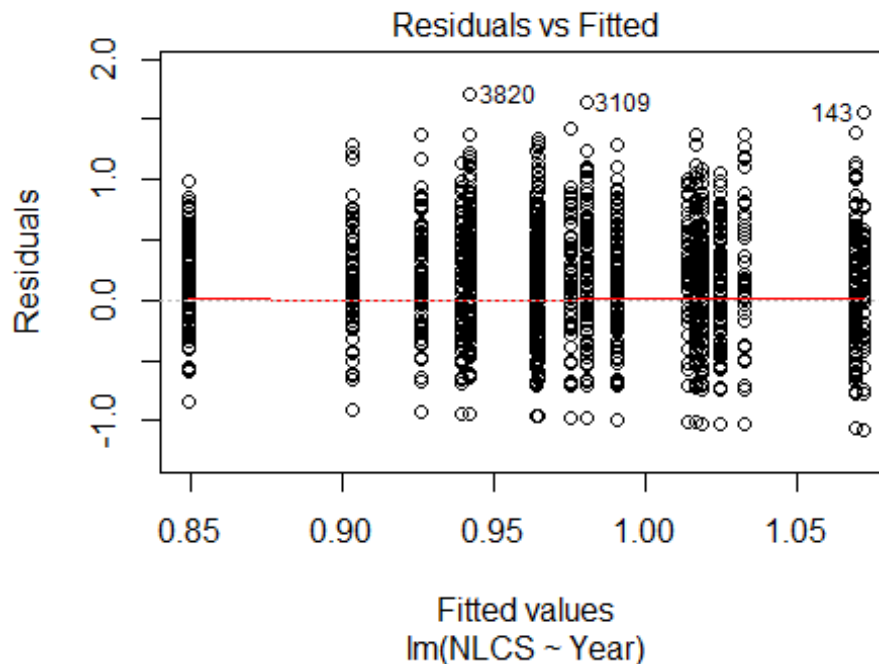
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4987 -0.3549 -0.0166 0.3616 2.2844
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4987 0.1443 10.38 < 2e-16 ***
## LastAuthorFemale1 0.0776 0.0329 2.36 0.01858 *
## Year1997 -0.1812 0.1632 -1.11 0.26705
## Year1998 -0.1418 0.1876 -0.76 0.44976
## Year1999 -1.3402 0.1500 -8.93 < 2e-16 ***
## Year2000 -0.3338 0.1664 -2.01 0.04510 *
## Year2001 -1.1422 0.1599 -7.14 1.4e-12 ***
## Year2002 -0.3736 0.1722 -2.17 0.03014 *
## Year2003 -0.4308 0.1615 -2.67 0.00772 **
## Year2004 -0.5256 0.1586 -3.31 0.00094 ***
## Year2005 -0.4520 0.1555 -2.91 0.00369 **
## Year2006 -0.4883 0.1562 -3.13 0.00181 **
```

```

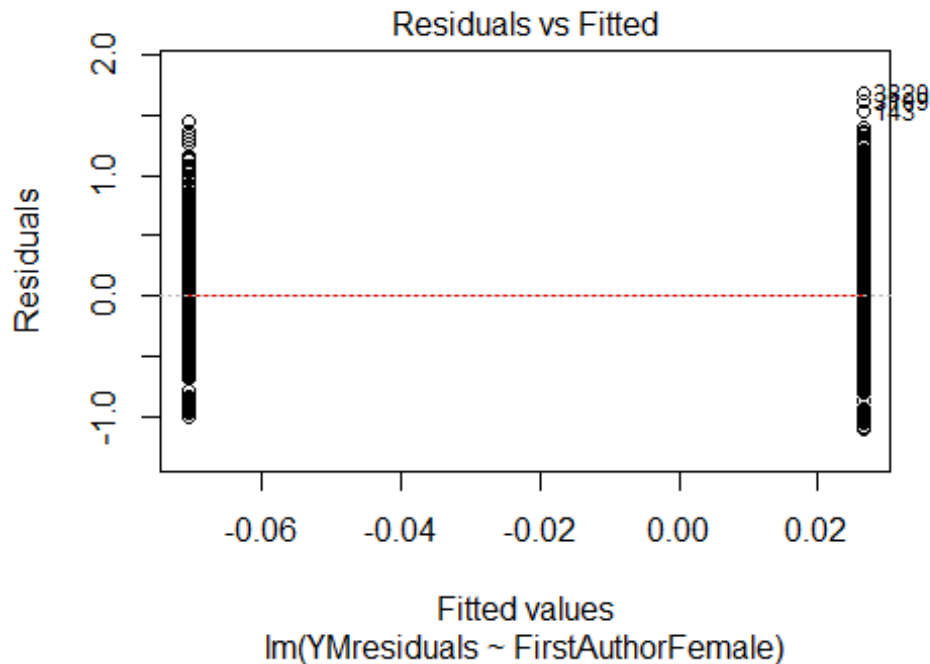
## Year2007          -0.4550      0.1530   -2.97  0.00298 **
## Year2008          -0.4730      0.1534   -3.08  0.00208 **
## Year2009          -0.3519      0.1523   -2.31  0.02103 *
## Year2010          -0.4099      0.1491   -2.75  0.00605 **
## Year2011          -0.3295      0.1502   -2.19  0.02840 *
## Year2012          -0.3755      0.1521   -2.47  0.01364 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.269, Adjusted R-squared:  0.261
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 233 is an outlier with |weight| = 0 ( < 6.4e-05);
## 100 weights are ~= 1. The remaining 1467 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0634 0.8450 0.9490 0.8870 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      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: 1568"
## [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
## 177 171 156 199 194 234 190 181 205 206 224 257 237 217 246
## 2011 2012
## 277 279
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 54 59 80 92 120 105 124 137 108 116 121 123 122 143

```

```
## 2011 2012
## 189 185
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 47 53 71 88 105 95 101 112 81 92 102 92 95 128
## 2011 2012
## 157 161
## [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
```

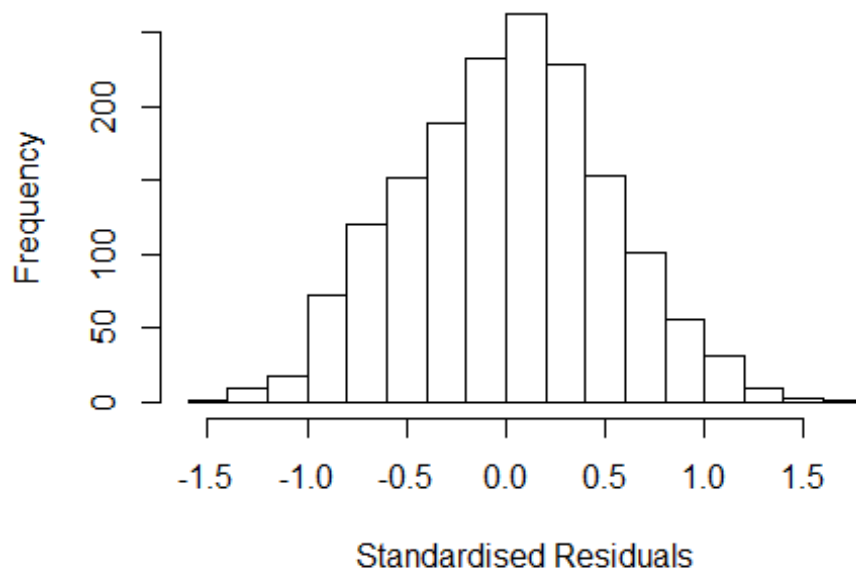


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.23, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 4.56926340818893"
## [1] "Male first author team size 2018 geometric mean: 4.06442644294368"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.33032583831044"
## [1] "Male last author team size 2018 geometric mean: 4.14507662073089"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9500, 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 1.087 1          1.043
## LastAuthorFemale  1.106 1          1.051
## UniqueAuthors    1.284 4          1.032
## Year              1.336 16         1.009
```

Residuals from first and last author and team size



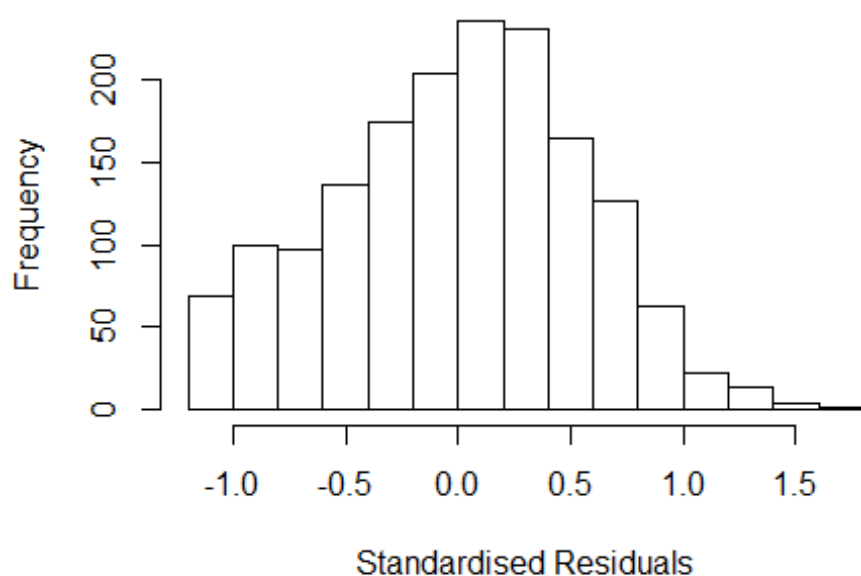
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4346 -0.3704  0.0214  0.3449  1.6148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7880    0.0771   10.22 < 2e-16 ***
## FirstAuthorFemale1 -0.0880    0.0304   -2.89  0.00384 **
## LastAuthorFemale1 -0.0604    0.0351   -1.72  0.08572 .
## UniqueAuthors2     0.2121    0.0581    3.65  0.00027 ***
## UniqueAuthors3     0.4180    0.0558    7.49  1.1e-13 ***
## UniqueAuthors4     0.4241    0.0555    7.65  3.5e-14 ***
## UniqueAuthors5     0.6466    0.0535   12.10 < 2e-16 ***
## Year1997         -0.0871    0.1133   -0.77  0.44224
## Year1998         -0.1085    0.1045   -1.04  0.29913
## Year1999         -0.1486    0.0857   -1.73  0.08332 .
```

```

## Year2000          -0.2343      0.0905    -2.59  0.00971 **
## Year2001          -0.2267      0.0834    -2.72  0.00662 **
## Year2002          -0.1722      0.0833    -2.07  0.03881 *
## Year2003          -0.2573      0.0807    -3.19  0.00146 **
## Year2004          -0.2243      0.0779    -2.88  0.00406 **
## Year2005          -0.1109      0.0868    -1.28  0.20137
## Year2006          -0.1061      0.0873    -1.22  0.22441
## Year2007          -0.1130      0.0869    -1.30  0.19384
## Year2008          -0.0863      0.0828    -1.04  0.29733
## Year2009          -0.1978      0.0877    -2.26  0.02417 *
## Year2010          -0.1881      0.0768    -2.45  0.01440 *
## Year2011          -0.2692      0.0798    -3.37  0.00076 ***
## Year2012          -0.2694      0.0802    -3.36  0.00080 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.16, Adjusted R-squared:  0.148
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1507 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.302  0.861  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          6.09e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1          1.026
## LastAuthorFemale  1.046 1          1.023
## Year              1.048 16          1.001

```


Residuals from first and last author



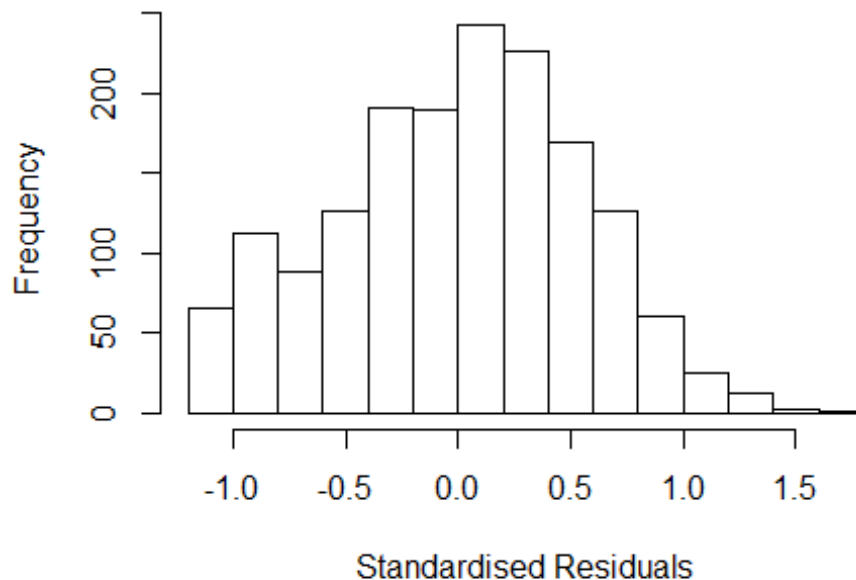
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1533 -0.3937 0.0354 0.3765 1.6706
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1170 0.0624 17.90 < 2e-16 ***
## FirstAuthorFemale1 -0.0861 0.0328 -2.62 0.00883 **
## LastAuthorFemale1 -0.1295 0.0391 -3.31 0.00095 ***
## Year1997 -0.0623 0.1310 -0.48 0.63426
## Year1998 -0.0842 0.1140 -0.74 0.46004
## Year1999 -0.0791 0.0875 -0.90 0.36607
## Year2000 -0.1691 0.0885 -1.91 0.05616 .
## Year2001 -0.1417 0.0816 -1.74 0.08260 .
## Year2002 -0.1063 0.0829 -1.28 0.19987
## Year2003 -0.2081 0.0803 -2.59 0.00964 **
## Year2004 -0.1054 0.0761 -1.38 0.16638
## Year2005 0.0124 0.0873 0.14 0.88680
```

```

## Year2006          -0.0189      0.0918   -0.21  0.83667
## Year2007          -0.0208      0.0906   -0.23  0.81847
## Year2008           0.0363      0.0835    0.43  0.66392
## Year2009          -0.0315      0.0920   -0.34  0.73224
## Year2010          -0.0703      0.0772   -0.91  0.36251
## Year2011          -0.1406      0.0816   -1.72  0.08512 .
## Year2012          -0.0956      0.0791   -1.21  0.22680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0265, Adjusted R-squared:  0.0157
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1512 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.368  0.868  0.951  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      6.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



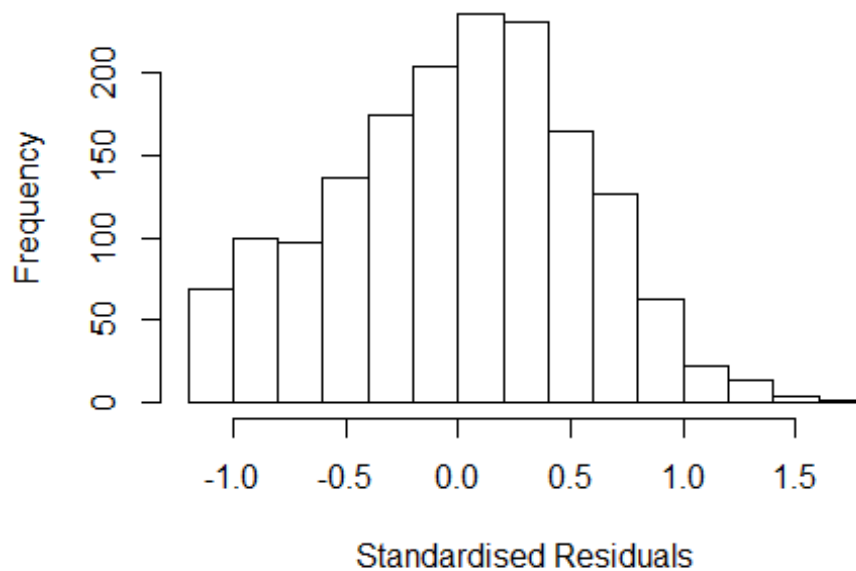
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1279 -0.3802  0.0419  0.3818  1.6905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0976     0.0616   17.81 < 2e-16 ***
## FirstAuthorFemale1 -0.1099     0.0330   -3.33  0.00089 ***
## Year1997          -0.0571     0.1293   -0.44  0.65911
## Year1998          -0.0838     0.1129   -0.74  0.45800
## Year1999          -0.0769     0.0864   -0.89  0.37348
## Year2000          -0.1568     0.0884   -1.77  0.07617 .
## Year2001          -0.1315     0.0815   -1.61  0.10700
## Year2002          -0.0995     0.0819   -1.21  0.22457
## Year2003          -0.2076     0.0796   -2.61  0.00920 **
## Year2004          -0.1033     0.0758   -1.36  0.17335
## Year2005           0.0166     0.0864    0.19  0.84793
## Year2006          -0.0190     0.0914   -0.21  0.83516
```

```

## Year2007          -0.0212      0.0894   -0.24  0.81291
## Year2008           0.0304      0.0834    0.36  0.71581
## Year2009          -0.0244      0.0922   -0.26  0.79150
## Year2010          -0.0709      0.0770   -0.92  0.35781
## Year2011          -0.1411      0.0815   -1.73  0.08371 .
## Year2012          -0.0928      0.0790   -1.17  0.24019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0192, Adjusted R-squared:  0.00888
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 126 weights are ~= 1. The remaining 1515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.357  0.868  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      6.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



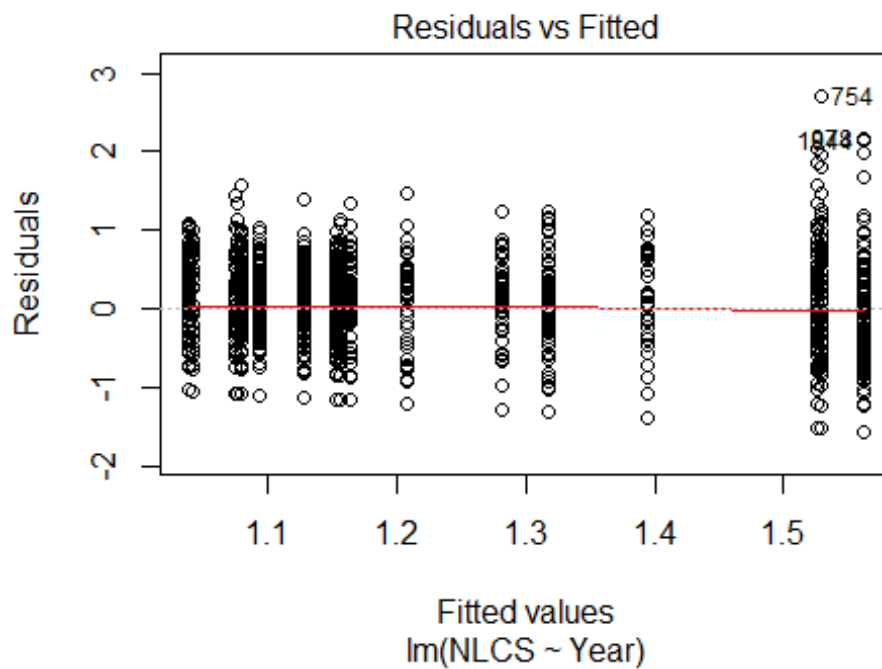
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1314 -0.3848  0.0272  0.3820  1.6953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10604    0.06202   17.83 < 2e-16 ***
## LastAuthorFemale1 -0.15088    0.03928   -3.84  0.00013 ***
## Year1997        -0.06547    0.13156   -0.50  0.61882
## Year1998        -0.08687    0.11427   -0.76  0.44724
## Year1999        -0.08407    0.08788   -0.96  0.33890
## Year2000        -0.17619    0.08889   -1.98  0.04762 *
## Year2001        -0.14651    0.08110   -1.81  0.07102 .
## Year2002        -0.11533    0.08294   -1.39  0.16456
## Year2003        -0.21520    0.08037   -2.68  0.00749 **
## Year2004        -0.10968    0.07580   -1.45  0.14813
## Year2005         0.00661    0.08677    0.08  0.93930
## Year2006        -0.02234    0.09238   -0.24  0.80896
```

```

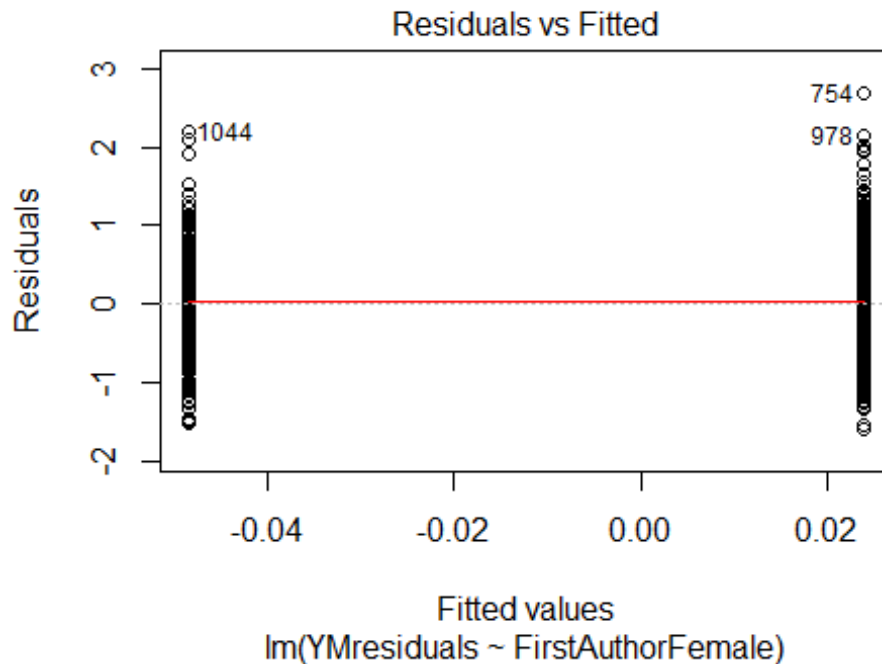
## Year2007          -0.03624      0.09042    -0.40  0.68860
## Year2008           0.02540      0.08352     0.30  0.76106
## Year2009          -0.04925      0.09184    -0.54  0.59184
## Year2010          -0.08069      0.07695    -1.05  0.29451
## Year2011          -0.15435      0.08145    -1.90  0.05826 .
## Year2012          -0.10876      0.07896    -1.38  0.16859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.012
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 142 weights are ~= 1. The remaining 1499 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.353  0.866  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      6.09e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1641"
## [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
##   63   74   61   91   68   75   81   93   96  115  133  144  166  180  177
## 2011 2012
##  231  203
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   50   40   69   34   38   58   76   85   95  113  119  138  139  140
## 2011 2012

```

```
## 179 159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 47 38 65 33 34 55 63 77 84 100 111 124 124 131
## 2011 2012
## 164 141
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 87, df = 16, p-value = 1e-11
```

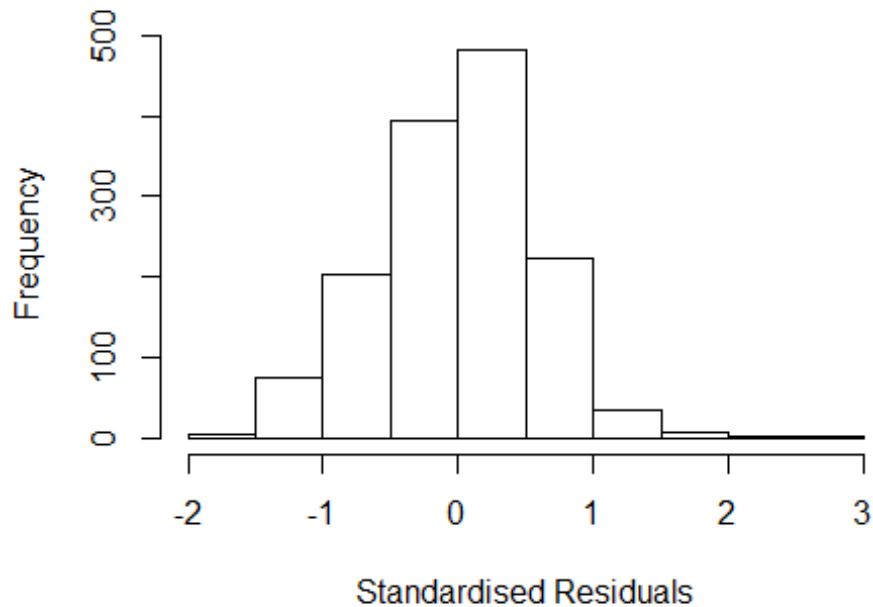


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.13, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 4.15686470804372"
## [1] "Male first author team size 2018 geometric mean: 3.54187363953176"
##
## 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] "Female last author team size 2018 geometric mean: 4.05362871783831"
## [1] "Male last author team size 2018 geometric mean: 3.91855996775381"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, 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 1.115 1      1.056
## LastAuthorFemale  1.114 1      1.056
## UniqueAuthors     1.267 4      1.030
## Year               1.344 16     1.009
```


Residuals from first and last author and team size



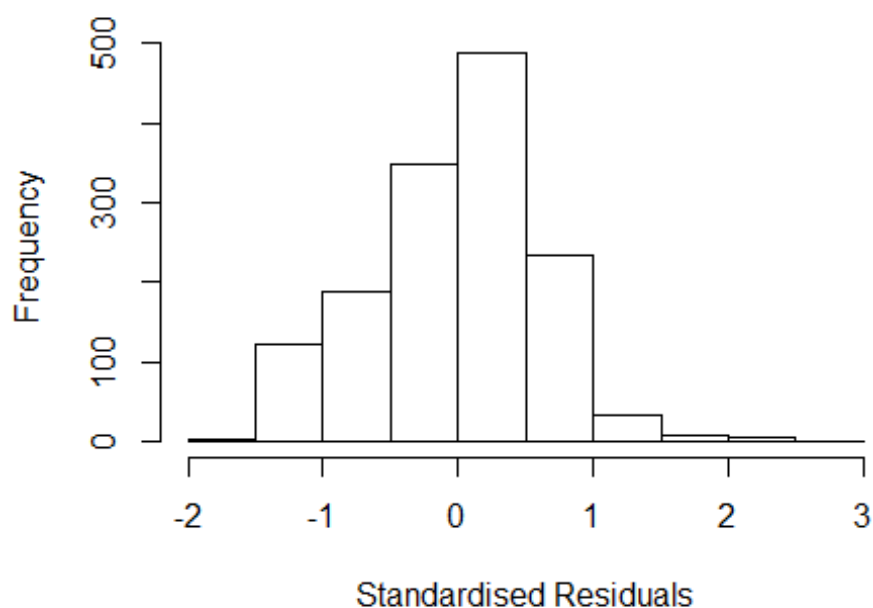
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754  2942691812 4.232 2004      2742      1      2.571
## 1044 33646158226 3.711 2006      2742      1      2.647
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7261 -0.3655  0.0255  0.3933  2.6468
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0169    0.1176   8.65 < 2e-16 ***
## FirstAuthorFemale1  0.0447    0.0338   1.32  0.18712
## LastAuthorFemale1  0.0298    0.0322   0.93  0.35432
## UniqueAuthors2     0.3283    0.0659   4.98 7.2e-07 ***
## UniqueAuthors3     0.4898    0.0632   7.76 1.7e-14 ***
## UniqueAuthors4     0.5874    0.0643   9.14 < 2e-16 ***
## UniqueAuthors5     0.6334    0.0632  10.02 < 2e-16 ***
## Year1997          -0.2237    0.1417  -1.58  0.11448
## Year1998          -0.3484    0.1314  -2.65  0.00809 **
```

```

## Year1999          -0.2757      0.1250    -2.21  0.02756 *
## Year2000          -0.3821      0.1548    -2.47  0.01372 *
## Year2001          -0.3265      0.1844    -1.77  0.07683 .
## Year2002          -0.1586      0.1301    -1.22  0.22323
## Year2003          -0.1102      0.1323    -0.83  0.40485
## Year2004          -0.0173      0.1414    -0.12  0.90246
## Year2005           0.0182      0.1279     0.14  0.88716
## Year2006           0.0473      0.1181     0.40  0.68872
## Year2007          -0.4407      0.1140    -3.87  0.00012 ***
## Year2008          -0.3558      0.1094    -3.25  0.00117 **
## Year2009          -0.3496      0.1076    -3.25  0.00118 **
## Year2010          -0.4006      0.1089    -3.68  0.00024 ***
## Year2011          -0.4332      0.1093    -3.97  7.7e-05 ***
## Year2012          -0.4430      0.1094    -4.05  5.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.554
## Multiple R-squared:  0.17,   Adjusted R-squared:  0.157
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 611 is an outlier with |weight| = 0 ( < 7e-05);
## 117 weights are ~= 1. The remaining 1311 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0004 0.8720 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          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 1.068 1          1.033
## LastAuthorFemale  1.068 1          1.033
## Year              1.097 16          1.003

```

Residuals from first and last author



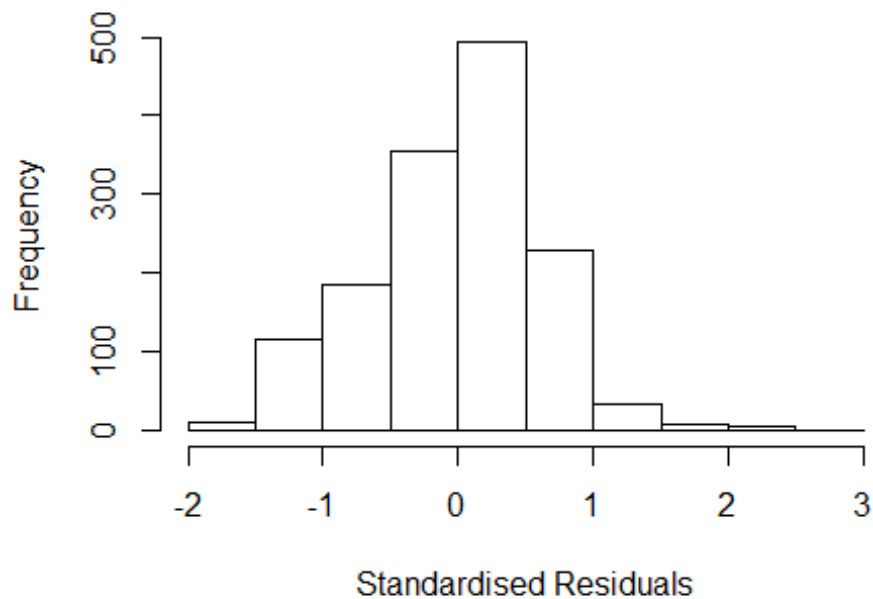
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 2942691812 4.232 2004      2742      1      2.737
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5486 -0.4130  0.0486  0.4024  2.7368
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3900     0.1159   11.99  <2e-16 ***
## FirstAuthorFemale1  0.0845     0.0361    2.34  0.0193 *
## LastAuthorFemale1 -0.0234     0.0336   -0.70  0.4863
## Year1997          -0.2034     0.1584   -1.28  0.1991
## Year1998          -0.3695     0.1426   -2.59  0.0097 **
## Year1999          -0.2668     0.1326   -2.01  0.0444 *
## Year2000          -0.4117     0.1697   -2.43  0.0154 *
## Year2001          -0.2650     0.1701   -1.56  0.1195
## Year2002          -0.0889     0.1393   -0.64  0.5237
## Year2003          -0.0834     0.1396   -0.60  0.5501
## Year2004           0.0440     0.1540    0.29  0.7750
## Year2005           0.0740     0.1388    0.53  0.5937
```

```

## Year2006          0.0965      0.1281      0.75      0.4513
## Year2007         -0.3846      0.1266     -3.04      0.0024 **
## Year2008         -0.2586      0.1184     -2.18      0.0291 *
## Year2009         -0.2673      0.1178     -2.27      0.0234 *
## Year2010         -0.3002      0.1191     -2.52      0.0118 *
## Year2011         -0.3443      0.1200     -2.87      0.0042 **
## Year2012         -0.3170      0.1201     -2.64      0.0084 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.0693, Adjusted R-squared:  0.0574
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 412 is an outlier with |weight| = 0 ( < 7e-05);
## 121 weights are ~= 1. The remaining 1307 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0978 0.8610 0.9440 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 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

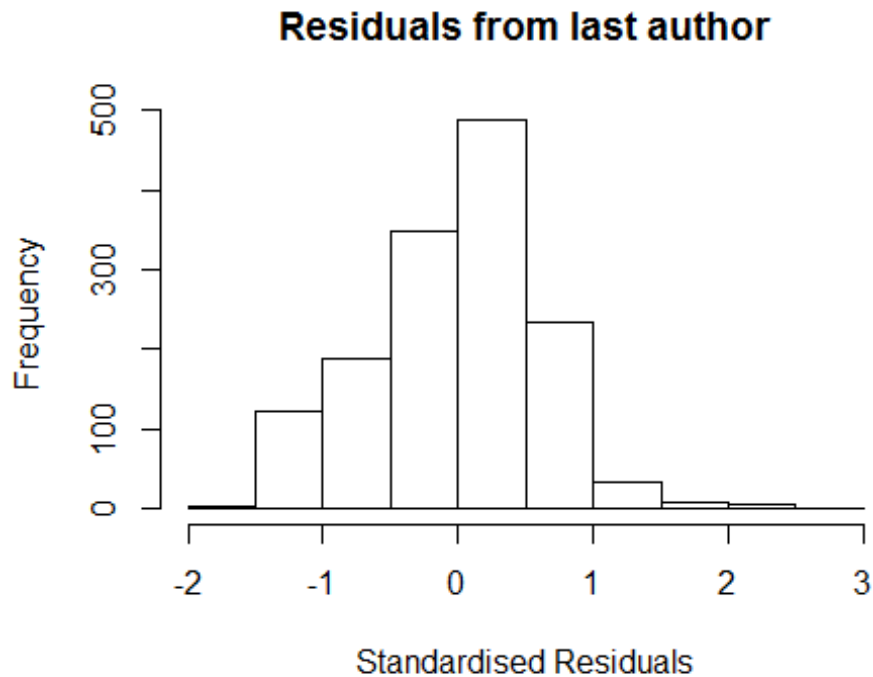


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 2942691812 4.232 2004      2742      1      2.737
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5554 -0.4120  0.0501  0.4082  2.7263
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3812     0.1148   12.03  <2e-16 ***
## FirstAuthorFemale1  0.0784     0.0359    2.19  0.0289 *
## Year1997         -0.1992     0.1586   -1.26  0.2093
## Year1998         -0.3733     0.1430   -2.61  0.0092 **
## Year1999         -0.2659     0.1328   -2.00  0.0455 *
## Year2000         -0.4068     0.1693   -2.40  0.0164 *
## Year2001         -0.2669     0.1703   -1.57  0.1173
## Year2002         -0.0880     0.1396   -0.63  0.5287
## Year2003         -0.0801     0.1394   -0.57  0.5655
## Year2004          0.0460     0.1543    0.30  0.7656
## Year2005          0.0749     0.1390    0.54  0.5900
## Year2006          0.0957     0.1284    0.75  0.4561
```

```

## Year2007          -0.3853      0.1269   -3.04   0.0024 **
## Year2008          -0.2599      0.1188   -2.19   0.0288 *
## Year2009          -0.2669      0.1181   -2.26   0.0239 *
## Year2010          -0.2987      0.1193   -2.50   0.0124 *
## Year2011          -0.3439      0.1202   -2.86   0.0043 **
## Year2012          -0.3159      0.1204   -2.62   0.0088 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.069, Adjusted R-squared:  0.0578
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 412 is an outlier with |weight| = 0 ( < 7e-05);
## 115 weights are ~= 1. The remaining 1313 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0934 0.8590 0.9450 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      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 1.05 1          1.024
## Year          1.05 16          1.002

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 754 2942691812 4.232 2004      2742      1      2.737
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5358 -0.4034  0.0502  0.3923  2.7494
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4338     0.1137  12.61  <2e-16 ***
## LastAuthorFemale1 -0.0044     0.0335  -0.13  0.8955
## Year1997         -0.2013     0.1609  -1.25  0.2111
## Year1998         -0.3811     0.1439  -2.65  0.0082 **
## Year1999         -0.2646     0.1336  -1.98  0.0479 *
## Year2000         -0.4033     0.1703  -2.37  0.0180 *
## Year2001         -0.2631     0.1699  -1.55  0.1218
## Year2002         -0.0985     0.1390  -0.71  0.4787
## Year2003         -0.0806     0.1407  -0.57  0.5670
## Year2004          0.0532     0.1552   0.34  0.7318
## Year2005          0.0796     0.1403   0.57  0.5704
## Year2006          0.1020     0.1290   0.79  0.4291
```

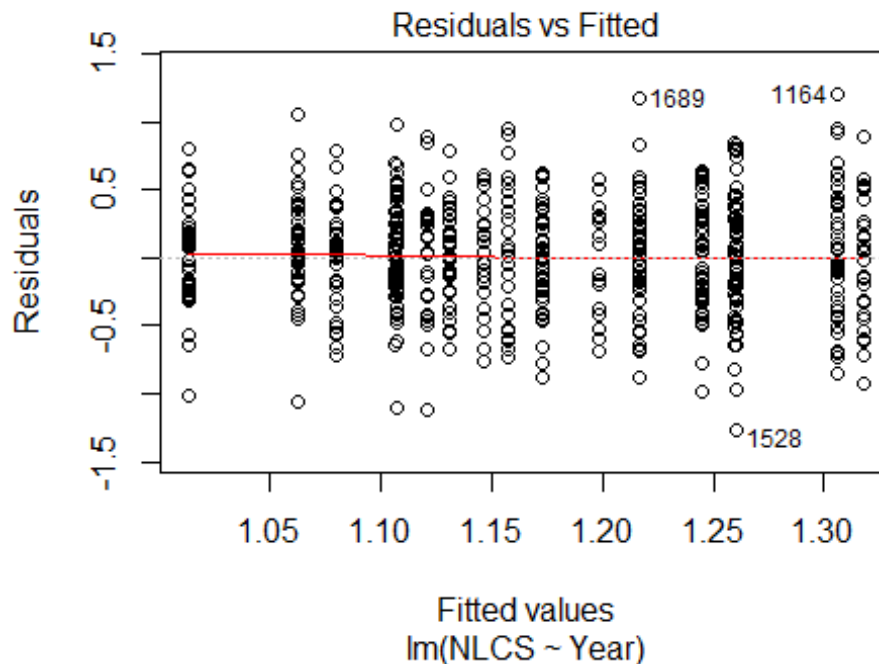
```

## Year2007          -0.3754      0.1273    -2.95    0.0033 **
## Year2008          -0.2622      0.1195    -2.19    0.0283 *
## Year2009          -0.2570      0.1190    -2.16    0.0309 *
## Year2010          -0.2965      0.1202    -2.47    0.0137 *
## Year2011          -0.3414      0.1210    -2.82    0.0048 **
## Year2012          -0.3110      0.1215    -2.56    0.0106 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.066, Adjusted R-squared:  0.0547
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 412 is an outlier with |weight| = 0 ( < 7e-05);
## 124 weights are ~= 1. The remaining 1304 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.856  0.944  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      7.00e-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: 1429"
## [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
##   67   71   59   60   64   90   93   78   72   61   78   81   92  102  105
## 2011 2012
##  112  102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   29   29   25   17   19   35   38   40   29   46   49   46   51   55

```



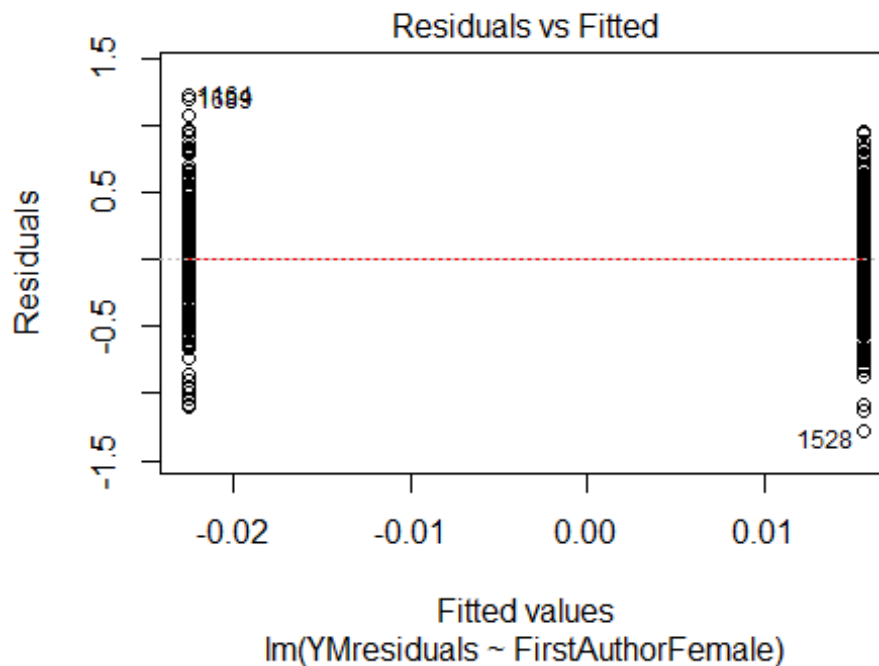
```
## 2011 2012
## 58 52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 24 27 22 12 18 32 19 28 23 41 39 40 49 51
## 2011 2012
## 56 47
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## 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 = 1.4, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 4.00947568573328"
## [1] "Male first author team size 2018 geometric mean: 4.3948106228412"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
```

```
## W = 260, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.92594951443832"
## [1] "Male last author team size 2018 geometric mean: 4.30679473055988"

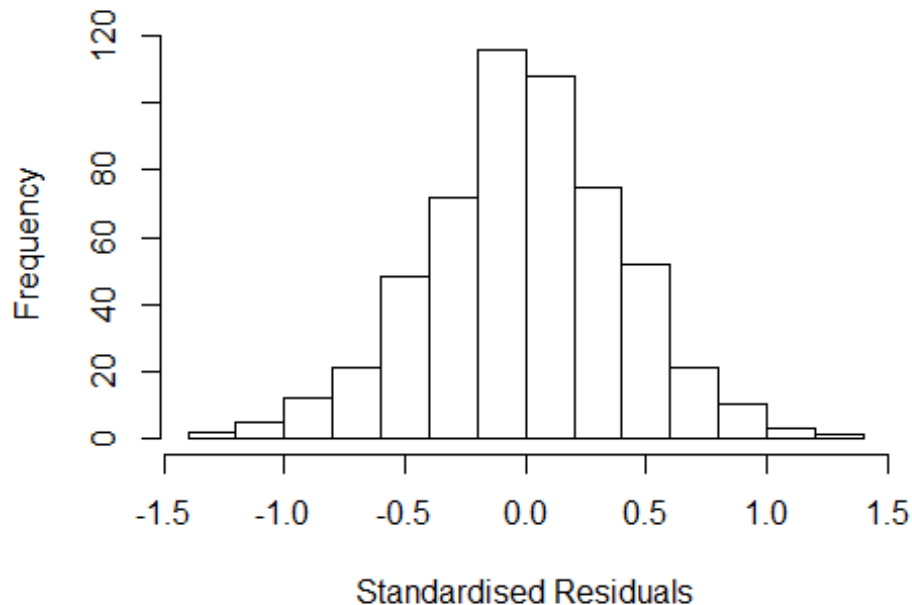
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 0.6
## 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	1.481	1	1.217
LastAuthorFemale	1.528	1	1.236
UniqueAuthors	2.179	4	1.102
Year	2.727	16	1.032

Residuals from first and last author and team size



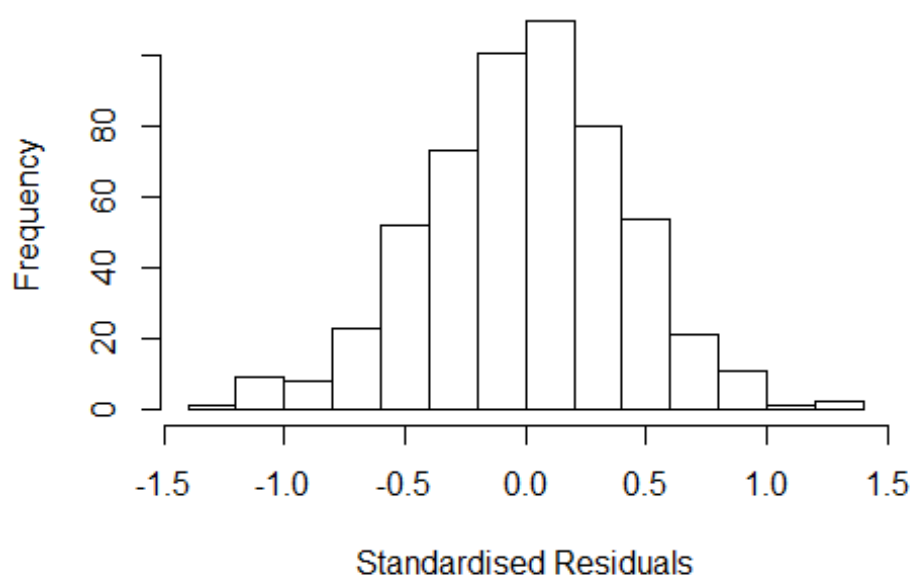
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34291 -0.27649 -0.00315  0.25787  1.26136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11770    0.13570   8.24 1.4e-15 ***
## FirstAuthorFemale1  0.04626    0.04247   1.09  0.2765
## LastAuthorFemale1 -0.04001    0.04246  -0.94  0.3466
## UniqueAuthors2     0.22420    0.07902   2.84  0.0047 **
## UniqueAuthors3     0.10765    0.07169   1.50  0.1338
## UniqueAuthors4     0.08849    0.07464   1.19  0.2363
## UniqueAuthors5     0.16455    0.06948   2.37  0.0182 *
## Year1997          0.09641    0.16035   0.60  0.5480
## Year1998         -0.08280    0.17673  -0.47  0.6396
## Year1999         -0.18029    0.16461  -1.10  0.2739
```

```

## Year2000      0.00368    0.17524    0.02    0.9833
## Year2001     -0.13366    0.13784   -0.97    0.3326
## Year2002     -0.20677    0.15041   -1.37    0.1698
## Year2003     -0.05973    0.15537   -0.38    0.7008
## Year2004     -0.21151    0.15159   -1.40    0.1635
## Year2005     -0.08490    0.15160   -0.56    0.5757
## Year2006     -0.16591    0.14915   -1.11    0.2665
## Year2007     -0.10179    0.15819   -0.64    0.5202
## Year2008      0.01829    0.17177    0.11    0.9153
## Year2009     -0.08361    0.14706   -0.57    0.5699
## Year2010     -0.03501    0.14803   -0.24    0.8131
## Year2011      0.01439    0.14736    0.10    0.9222
## Year2012     -0.03343    0.14769   -0.23    0.8210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0651, Adjusted R-squared:  0.0258
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 494 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.868  0.946  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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.418 1      1.191
## LastAuthorFemale  1.374 1      1.172
## Year              1.426 16      1.011

```

Residuals from first and last author



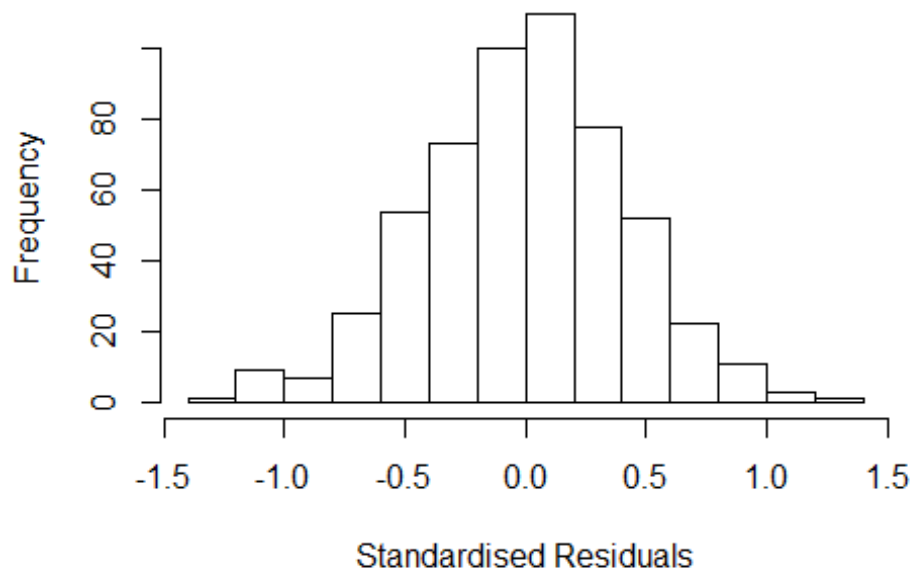
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31406 -0.27470  0.00867  0.27104  1.24211
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2089    0.1315   9.20  <2e-16 ***
## FirstAuthorFemale1  0.0471    0.0425   1.11    0.27
## LastAuthorFemale1 -0.0418    0.0413  -1.01    0.31
## Year1997          0.0997    0.1611   0.62    0.54
## Year1998         -0.0415    0.1827  -0.23    0.82
## Year1999         -0.1491    0.1672  -0.89    0.37
## Year2000          0.0291    0.1764   0.16    0.87
## Year2001         -0.1422    0.1406  -1.01    0.31
## Year2002         -0.1421    0.1478  -0.96    0.34
## Year2003         -0.0186    0.1544  -0.12    0.90
## Year2004         -0.1755    0.1483  -1.18    0.24
## Year2005         -0.0870    0.1512  -0.58    0.57
```

```

## Year2006          -0.1382      0.1497    -0.92      0.36
## Year2007          -0.0786      0.1588    -0.49      0.62
## Year2008           0.0540      0.1641      0.33      0.74
## Year2009          -0.0442      0.1451    -0.30      0.76
## Year2010           0.0139      0.1459      0.10      0.92
## Year2011           0.0580      0.1459      0.40      0.69
## Year2012           0.0123      0.1452      0.08      0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0411, Adjusted R-squared:  0.00834
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 505 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.276  0.873   0.952   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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.247 1      1.117
## Year              1.247 16      1.007

```

Residuals from first author



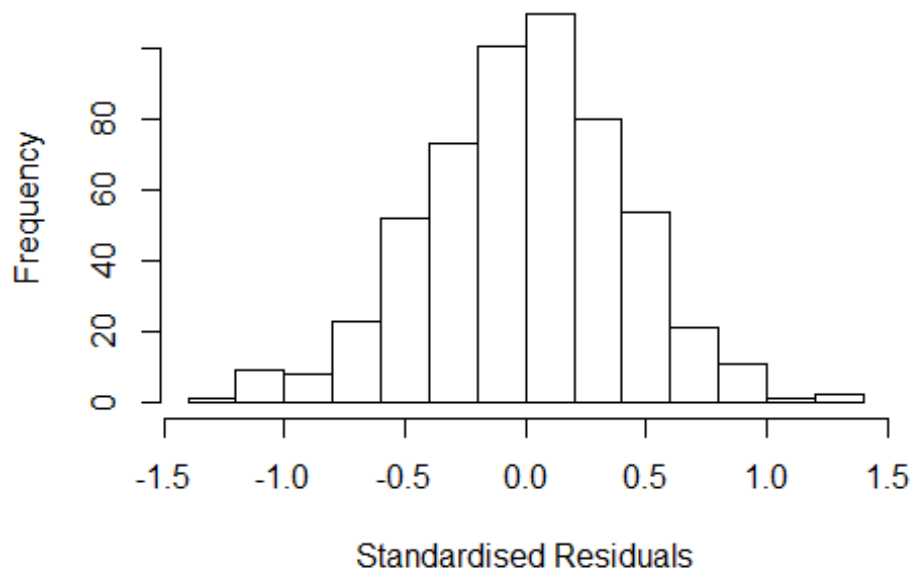
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2883 -0.2727 0.0147 0.2725 1.2530
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21e+00 1.30e-01 9.29 <2e-16 ***
## FirstAuthorFemale1 3.73e-02 4.02e-02 0.93 0.35
## Year1997 9.66e-02 1.60e-01 0.60 0.55
## Year1998 -4.22e-02 1.83e-01 -0.23 0.82
## Year1999 -1.53e-01 1.66e-01 -0.92 0.36
## Year2000 2.50e-02 1.75e-01 0.14 0.89
## Year2001 -1.45e-01 1.38e-01 -1.05 0.29
## Year2002 -1.43e-01 1.46e-01 -0.98 0.33
## Year2003 -1.59e-02 1.54e-01 -0.10 0.92
## Year2004 -1.87e-01 1.46e-01 -1.28 0.20
## Year2005 -8.93e-02 1.49e-01 -0.60 0.55
## Year2006 -1.44e-01 1.47e-01 -0.98 0.33
```

```

## Year2007          -9.09e-02   1.55e-01   -0.59    0.56
## Year2008          4.64e-02   1.62e-01    0.29    0.77
## Year2009         -5.27e-02   1.43e-01   -0.37    0.71
## Year2010          2.61e-04   1.43e-01    0.00    1.00
## Year2011          4.55e-02   1.43e-01    0.32    0.75
## Year2012          6.02e-05   1.43e-01    0.00    1.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0387, Adjusted R-squared:  0.00773
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 506 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.296  0.874  0.952  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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.196 1      1.094
## Year              1.196 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
## -1.2949 -0.2710  0.0142  0.2584  1.2126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22460    0.13443   9.11   <2e-16 ***
## LastAuthorFemale1 -0.03090    0.03890  -0.79    0.43
## Year1997         0.09746    0.16558   0.59    0.56
## Year1998        -0.03917    0.18510  -0.21    0.83
## Year1999        -0.13270    0.16781  -0.79    0.43
## Year2000         0.03395    0.17918   0.19    0.85
## Year2001        -0.14455    0.14362  -1.01    0.31
## Year2002        -0.14279    0.15045  -0.95    0.34
## Year2003        -0.00397    0.15494  -0.03    0.98
## Year2004        -0.17294    0.15006  -1.15    0.25
## Year2005        -0.07661    0.15300  -0.50    0.62
## Year2006        -0.12724    0.15083  -0.84    0.40
```

```

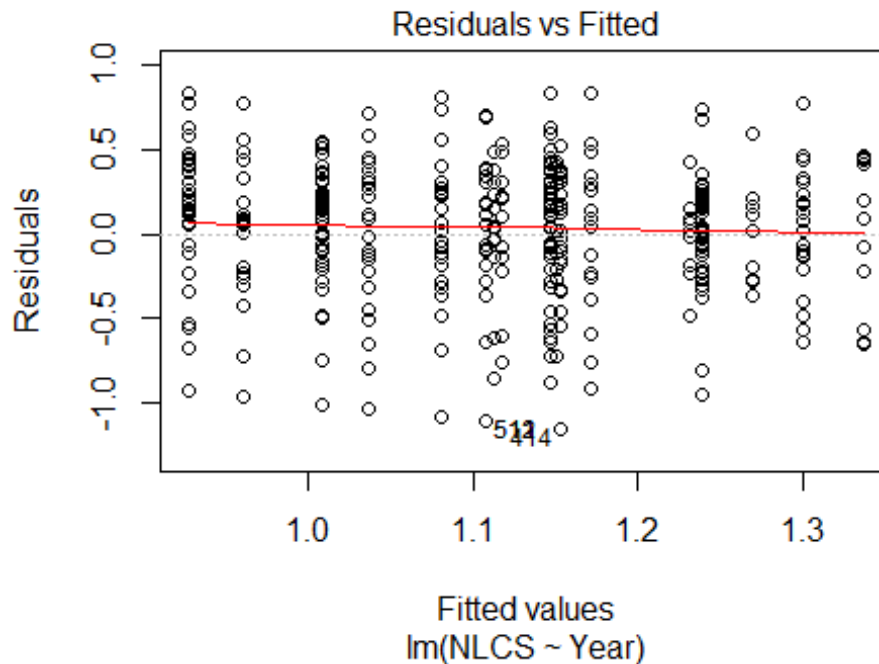
## Year2007          -0.06512      0.15911    -0.41      0.68
## Year2008           0.06777      0.16365      0.41      0.68
## Year2009          -0.02862      0.14578    -0.20      0.84
## Year2010           0.02421      0.14725      0.16      0.87
## Year2011           0.07030      0.14693      0.48      0.63
## Year2012           0.02008      0.14665      0.14      0.89
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.00735
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 506 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.289  0.875   0.952   0.905   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      1.83e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 546"
## [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
##    1    3    1
##
## 1996 1997 2006
##    0    3    1
##
## 1996 1997 2006
##    0    3    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: 4"
## [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
## 21 36 31 23 37 44 28 34 41 42 46 46 67 58 58
## 2011 2012
## 64 70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 17 13 9 12 10 21 12 22 22 25 23 42 32 46
## 2011 2012
## 46 45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 15 11 9 11 9 19 9 19 18 23 19 33 29 41
## 2011 2012
## 40 36
## [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

```



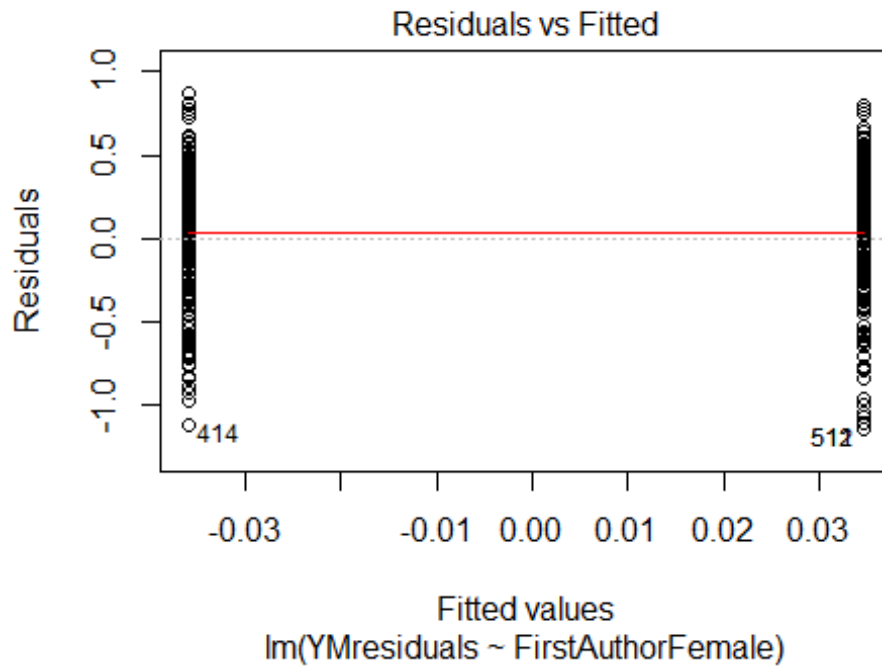
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07

## [1] "Female first author team size 2018 geometric mean: 6.27753767256866"
## [1] "Male first author team size 2018 geometric mean: 5.01130942080297"

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

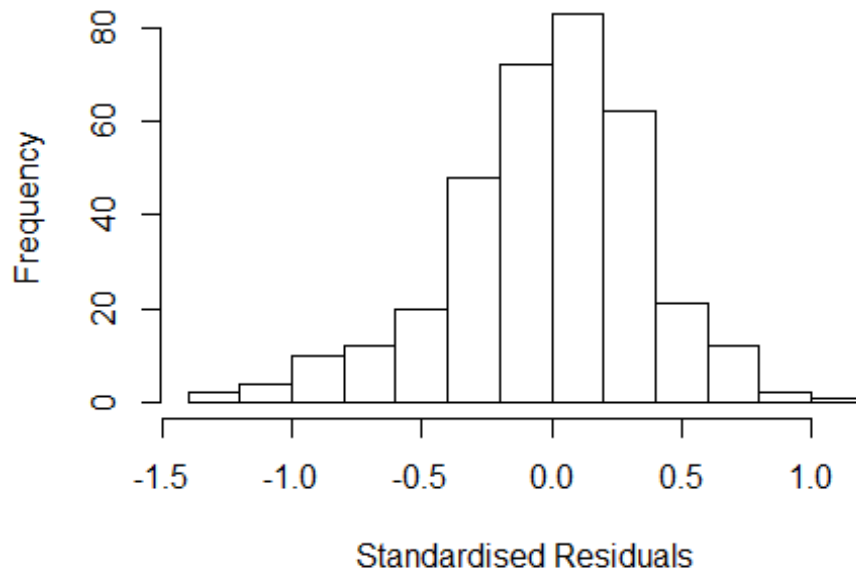
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.16637564176771"
## [1] "Male last author team size 2018 geometric mean: 5.7701516635039"

## 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 = 0.5
## 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 1.441 1          1.201
## LastAuthorFemale  1.422 1          1.192
## UniqueAuthors    3.225 4          1.158
## Year              4.997 16         1.052
```

Residuals from first and last author and team size



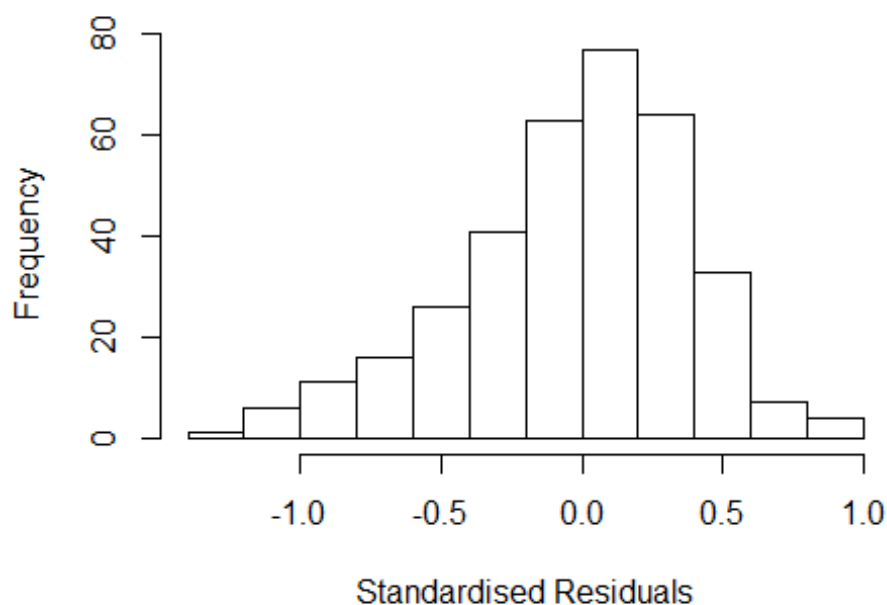
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2325 -0.2286 0.0203 0.2173 1.1700
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95868 0.16332 5.87 1.1e-08 ***
## FirstAuthorFemale1 0.02183 0.04519 0.48 0.6294
## LastAuthorFemale1 0.06457 0.05008 1.29 0.1982
## UniqueAuthors2 -0.00823 0.14154 -0.06 0.9537
## UniqueAuthors3 0.23258 0.13374 1.74 0.0830 .
## UniqueAuthors4 0.35013 0.12654 2.77 0.0060 **
## UniqueAuthors5 0.40381 0.13013 3.10 0.0021 **
## Year1997 -0.02961 0.15940 -0.19 0.8528
## Year1998 -0.09392 0.15282 -0.61 0.5392
## Year1999 0.05004 0.15326 0.33 0.7442
```

```

## Year2000      0.11297    0.16313    0.69    0.4891
## Year2001      0.04182    0.13726    0.30    0.7608
## Year2002     -0.04449    0.15246   -0.29    0.7706
## Year2003     -0.09959    0.17051   -0.58    0.5596
## Year2004     -0.28218    0.13920   -2.03    0.0435 *
## Year2005     -0.12871    0.15619   -0.82    0.4105
## Year2006     -0.21318    0.13856   -1.54    0.1249
## Year2007     -0.15185    0.14268   -1.06    0.2880
## Year2008     -0.40857    0.14129   -2.89    0.0041 **
## Year2009     -0.19590    0.12770   -1.53    0.1260
## Year2010     -0.25560    0.12411   -2.06    0.0402 *
## Year2011     -0.17101    0.12818   -1.33    0.1831
## Year2012     -0.09701    0.12267   -0.79    0.4296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.225, Adjusted R-squared:  0.172
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 320 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.163  0.866  0.953   0.887   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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.285 1      1.134
## LastAuthorFemale  1.264 1      1.124
## Year              1.622 16      1.015

```

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
## -1.208 -0.245  0.035  0.254  0.891
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1976    0.1280   9.36  <2e-16 ***
## FirstAuthorFemale1  0.0602    0.0479   1.26   0.210
## LastAuthorFemale1  0.0571    0.0499   1.14   0.253
## Year1997          -0.0823    0.1951  -0.42   0.674
## Year1998          -0.1397    0.1661  -0.84   0.401
## Year1999           0.0377    0.1644   0.23   0.819
## Year2000           0.1163    0.1968   0.59   0.555
## Year2001           0.0118    0.1564   0.08   0.940
## Year2002           0.0491    0.1569   0.31   0.754
## Year2003          -0.1183    0.1901  -0.62   0.534
## Year2004          -0.2952    0.1572  -1.88   0.061 .
## Year2005          -0.0462    0.1551  -0.30   0.766
```

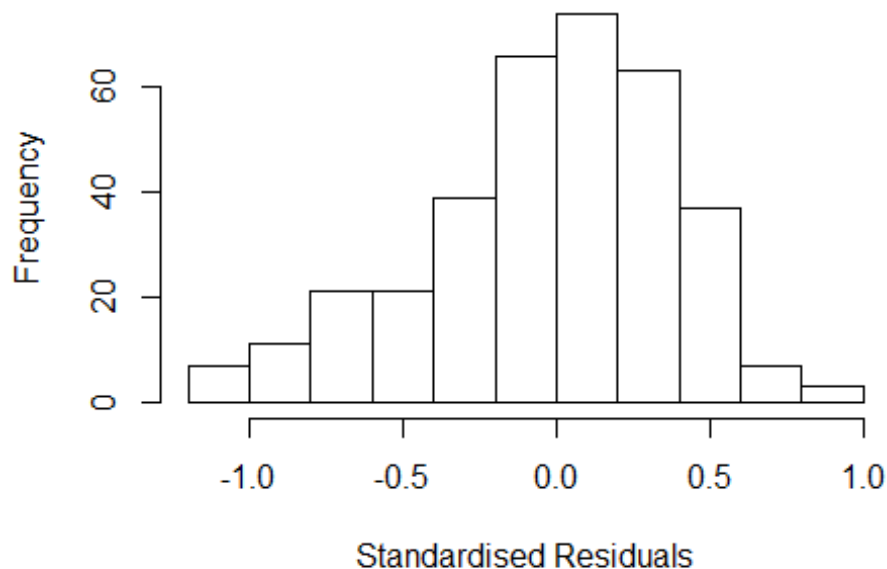


```

## Year2006          -0.1803      0.1561    -1.16     0.249
## Year2007          -0.0942      0.1643    -0.57     0.567
## Year2008          -0.3648      0.1698    -2.15     0.032 *
## Year2009          -0.1431      0.1511    -0.95     0.344
## Year2010          -0.1969      0.1405    -1.40     0.162
## Year2011          -0.1326      0.1475    -0.90     0.369
## Year2012          -0.0262      0.1418    -0.18     0.854
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0884, Adjusted R-squared:  0.0387
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 324 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.280  0.860  0.955  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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



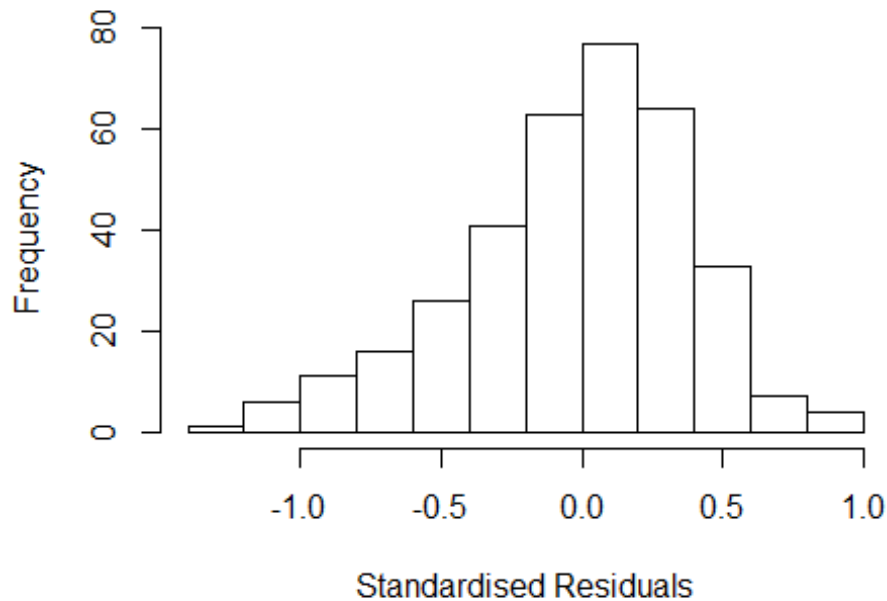
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1836 -0.2405 0.0344 0.2526 0.8950
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2041 0.1320 9.12 <2e-16 ***
## FirstAuthorFemale1 0.0665 0.0483 1.38 0.170
## Year1997 -0.0931 0.1976 -0.47 0.638
## Year1998 -0.1413 0.1697 -0.83 0.406
## Year1999 0.0286 0.1678 0.17 0.865
## Year2000 0.1152 0.1986 0.58 0.562
## Year2001 0.0189 0.1560 0.12 0.904
## Year2002 0.0609 0.1614 0.38 0.706
## Year2003 -0.1142 0.1949 -0.59 0.558
## Year2004 -0.2930 0.1607 -1.82 0.069 .
## Year2005 -0.0437 0.1576 -0.28 0.782
## Year2006 -0.1556 0.1571 -0.99 0.323
```

```

## Year2007          -0.0870      0.1677   -0.52    0.604
## Year2008          -0.3512      0.1727   -2.03    0.043 *
## Year2009          -0.1343      0.1551   -0.87    0.387
## Year2010          -0.1813      0.1434   -1.26    0.207
## Year2011          -0.1227      0.1507   -0.81    0.416
## Year2012          -0.0198      0.1458   -0.14    0.892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0835, Adjusted R-squared:  0.0364
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 325 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.302  0.857   0.954   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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.282 1          1.132
## Year            1.282 16          1.008

```

Residuals from last author



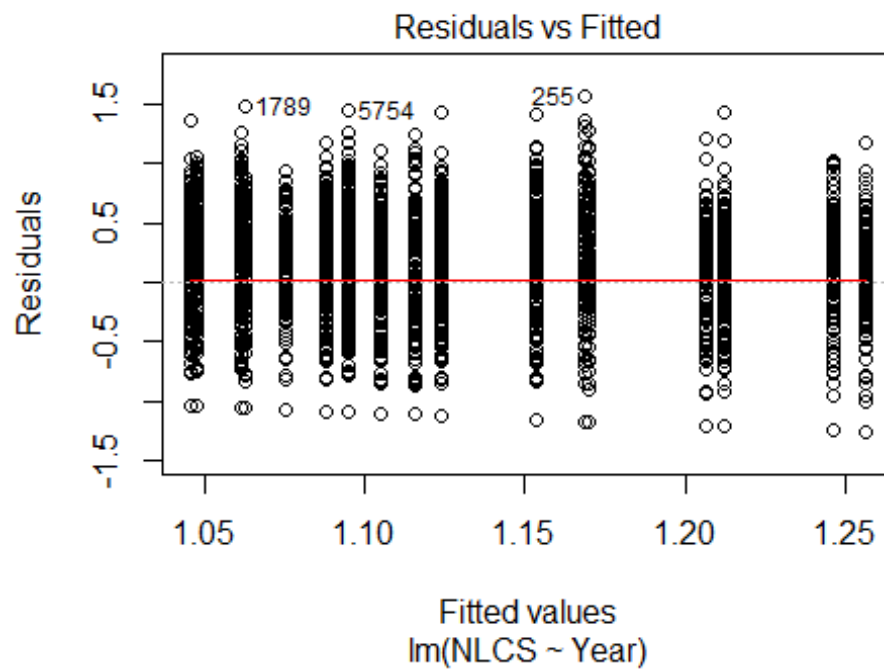
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2444 -0.2524 0.0339 0.2636 0.8496
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20489 0.12274 9.82 <2e-16 ***
## LastAuthorFemale1 0.06501 0.05060 1.28 0.200
## Year1997 -0.04854 0.18693 -0.26 0.795
## Year1998 -0.13097 0.15975 -0.82 0.413
## Year1999 0.05440 0.15918 0.34 0.733
## Year2000 0.12736 0.19858 0.64 0.522
## Year2001 0.03980 0.14948 0.27 0.790
## Year2002 0.08124 0.14985 0.54 0.588
## Year2003 -0.10415 0.18911 -0.55 0.582
## Year2004 -0.26149 0.14968 -1.75 0.082 .
## Year2005 -0.02547 0.14966 -0.17 0.865
## Year2006 -0.17072 0.15358 -1.11 0.267
```

```

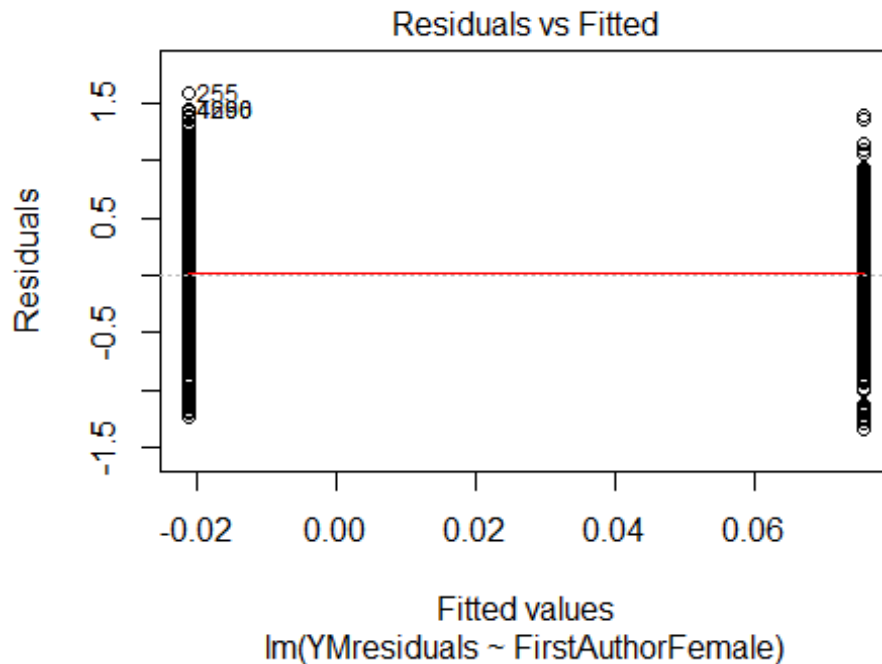
## Year2007          -0.07566      0.16071    -0.47      0.638
## Year2008          -0.33897      0.16314    -2.08      0.038 *
## Year2009          -0.10899      0.14225    -0.77      0.444
## Year2010          -0.18244      0.13548    -1.35      0.179
## Year2011          -0.11172      0.14185    -0.79      0.432
## Year2012          -0.00347      0.13454    -0.03      0.979
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.0836, Adjusted R-squared:  0.0365
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 327 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.247  0.855   0.953   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      2.87e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 349"
## [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
## 349 337 336 312 286 331 325 242 301 348 385 364 422 419 427
## 2011 2012
## 429 398
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 139 143 144 127 168 206 164 209 242 260 227 294 274 308
## 2011 2012

```

```
## 310 289
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 127 129 134 111 154 176 137 183 196 221 203 243 244 274
## 2011 2012
## 279 272
## [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
```

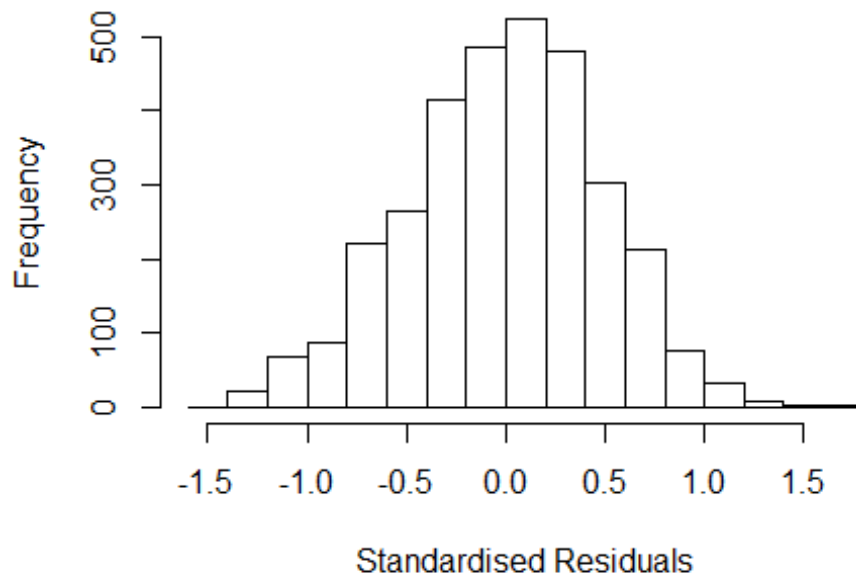


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.31371350250661"
## [1] "Male first author team size 2018 geometric mean: 4.25137083966946"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.40488985621193"
## [1] "Male last author team size 2018 geometric mean: 4.24397189198821"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, 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 1.098 1      1.048
## LastAuthorFemale  1.033 1      1.016
## UniqueAuthors    1.171 4      1.020
## 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
## -1.4504 -0.3330 0.0138 0.3246 1.7176
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7874 0.0686 11.48 < 2e-16 ***
## FirstAuthorFemale1 0.0622 0.0204 3.05 0.0023 **
## LastAuthorFemale1 0.0827 0.0258 3.20 0.0014 **
## UniqueAuthors2 0.3077 0.0489 6.29 3.6e-10 ***
## UniqueAuthors3 0.3576 0.0477 7.50 8.1e-14 ***
## UniqueAuthors4 0.4182 0.0484 8.64 < 2e-16 ***
## UniqueAuthors5 0.5902 0.0458 12.88 < 2e-16 ***
## Year1997 0.0106 0.0705 0.15 0.8810
## Year1998 0.0754 0.0679 1.11 0.2671
## Year1999 0.0807 0.0711 1.13 0.2567
```

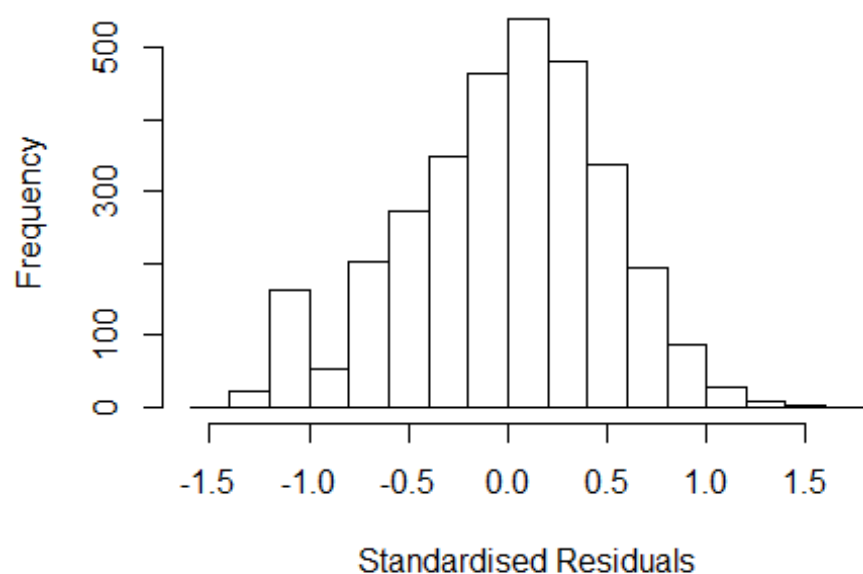


```

## Year2000          -0.0283      0.0710   -0.40   0.6899
## Year2001          -0.1100      0.0657   -1.67   0.0942 .
## Year2002          -0.1240      0.0656   -1.89   0.0587 .
## Year2003          -0.1632      0.0651   -2.51   0.0122 *
## Year2004          -0.1155      0.0666   -1.73   0.0830 .
## Year2005          -0.1176      0.0660   -1.78   0.0751 .
## Year2006          -0.1675      0.0627   -2.67   0.0076 **
## Year2007          -0.0395      0.0629   -0.63   0.5299
## Year2008          -0.1148      0.0633   -1.81   0.0699 .
## Year2009          -0.1596      0.0659   -2.42   0.0154 *
## Year2010          -0.1422      0.0632   -2.25   0.0245 *
## Year2011          -0.0892      0.0635   -1.40   0.1604
## Year2012          -0.1875      0.0641   -2.92   0.0035 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.122, Adjusted R-squared:  0.116
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 286 weights are ~= 1. The remaining 2916 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.172  0.862  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          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 1.080 1          1.039
## LastAuthorFemale  1.023 1          1.011
## Year              1.093 16          1.003

```

Residuals from first and last author



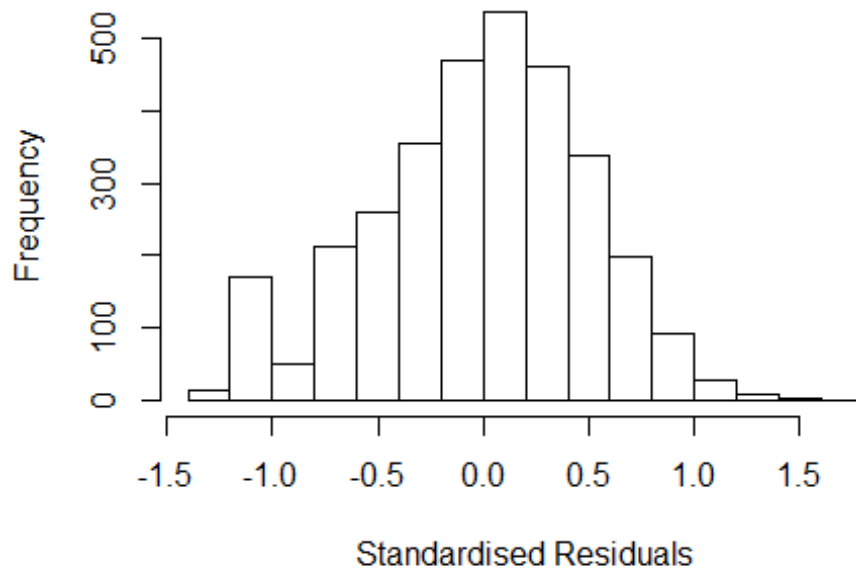
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4231 -0.3357  0.0259  0.3352  1.6072
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1278    0.0567   19.88 < 2e-16 ***
## FirstAuthorFemale1 0.1005    0.0211    4.75 2.1e-06 ***
## LastAuthorFemale1  0.0869    0.0272    3.19 0.0014 **
## Year1997          0.0216    0.0727    0.30 0.7663
## Year1998          0.1079    0.0698    1.55 0.1221
## Year1999          0.1126    0.0725    1.55 0.1205
## Year2000          0.0330    0.0740    0.45 0.6555
## Year2001         -0.0620    0.0694   -0.89 0.3720
## Year2002         -0.0608    0.0681   -0.89 0.3722
## Year2003         -0.1057    0.0679   -1.56 0.1197
## Year2004         -0.0466    0.0687   -0.68 0.4979
## Year2005         -0.0493    0.0677   -0.73 0.4664
```

```

## Year2006          -0.1069      0.0658   -1.63   0.1042
## Year2007           0.0577      0.0648    0.89   0.3727
## Year2008          -0.0256      0.0666   -0.39   0.7002
## Year2009          -0.0712      0.0687   -1.04   0.3002
## Year2010          -0.0527      0.0671   -0.79   0.4324
## Year2011           0.0121      0.0660    0.18   0.8550
## Year2012          -0.1029      0.0669   -1.54   0.1243
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0255, Adjusted R-squared:  0.02
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 265 weights are ~= 1. The remaining 2937 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.268  0.865  0.948  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      3.12e-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 1.07 1      1.034
## Year              1.07 16      1.002

```

Residuals from first author



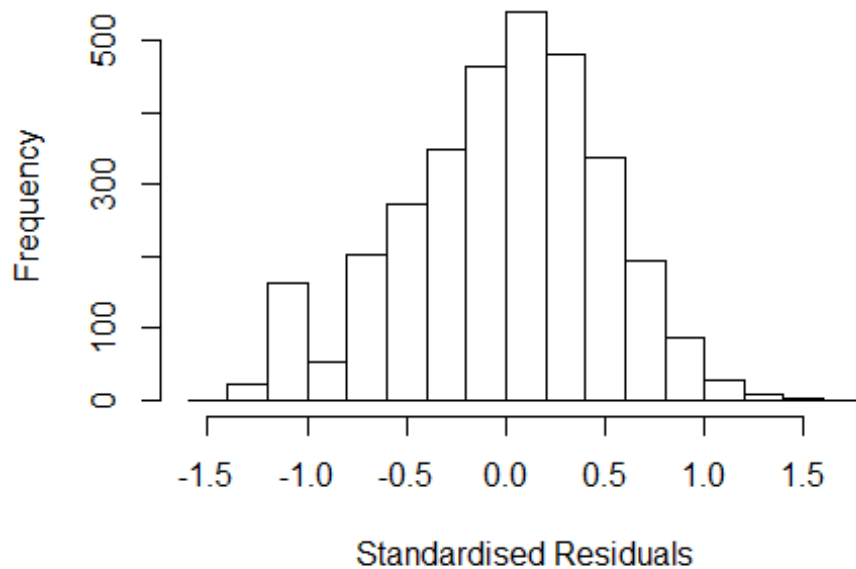
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3556 -0.3375 0.0275 0.3425 1.6003
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1347 0.0569 19.94 < 2e-16 ***
## FirstAuthorFemale1 0.1105 0.0211 5.23 1.8e-07 ***
## Year1997 0.0245 0.0730 0.34 0.74
## Year1998 0.1104 0.0700 1.58 0.11
## Year1999 0.1142 0.0727 1.57 0.12
## Year2000 0.0359 0.0743 0.48 0.63
## Year2001 -0.0584 0.0695 -0.84 0.40
## Year2002 -0.0578 0.0682 -0.85 0.40
## Year2003 -0.1018 0.0681 -1.49 0.14
## Year2004 -0.0450 0.0688 -0.65 0.51
## Year2005 -0.0475 0.0680 -0.70 0.49
## Year2006 -0.1061 0.0660 -1.61 0.11
```

```

## Year2007          0.0571      0.0651      0.88      0.38
## Year2008          -0.0227     0.0668     -0.34     0.73
## Year2009          -0.0699     0.0689     -1.01     0.31
## Year2010          -0.0562     0.0672     -0.84     0.40
## Year2011           0.0135     0.0661      0.20     0.84
## Year2012          -0.0980     0.0671     -1.46     0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.017
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 263 weights are ~= 1. The remaining 2939 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.278  0.864  0.950  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      3.12e-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 1.015 1          1.008
## Year              1.015 16          1.000

```

Residuals from last author



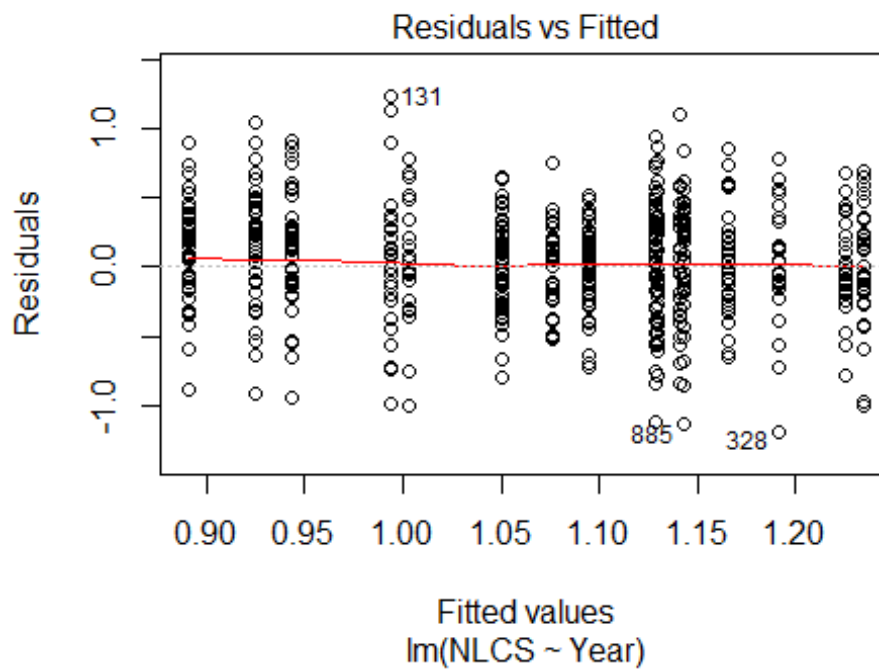
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3558 -0.3469 0.0314 0.3409 1.5983
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1367 0.0564 20.15 < 2e-16 ***
## LastAuthorFemale1 0.1049 0.0271 3.87 0.00011 ***
## Year1997 0.0271 0.0727 0.37 0.70895
## Year1998 0.1142 0.0692 1.65 0.09916 .
## Year1999 0.1171 0.0722 1.62 0.10501
## Year2000 0.0439 0.0736 0.60 0.55038
## Year2001 -0.0631 0.0692 -0.91 0.36151
## Year2002 -0.0566 0.0680 -0.83 0.40497
## Year2003 -0.0997 0.0676 -1.48 0.14027
## Year2004 -0.0361 0.0684 -0.53 0.59822
## Year2005 -0.0368 0.0674 -0.55 0.58528
## Year2006 -0.0967 0.0655 -1.47 0.14042
```

```

## Year2007          0.0727      0.0645      1.13  0.25954
## Year2008          -0.0116     0.0663     -0.17  0.86158
## Year2009          -0.0601     0.0686     -0.88  0.38145
## Year2010          -0.0329     0.0666     -0.49  0.62110
## Year2011           0.0338     0.0654      0.52  0.60515
## Year2012          -0.0804     0.0665     -1.21  0.22655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.495
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.0139
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 250 weights are ~= 1. The remaining 2952 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.276  0.865  0.950  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      3.12e-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 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
##   44   39   53   40   62   60   56   46   57   63   73   54   72   63   66
## 2011 2012
##   79   80
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   13   27   24   32   27   34   33   36   54   54   30   47   41   46
## 2011 2012

```

```
## 56 55
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 12 23 22 30 21 32 29 33 45 42 29 36 39 44
## 2011 2012
## 49 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 = 64, df = 16, p-value = 1e-07
```

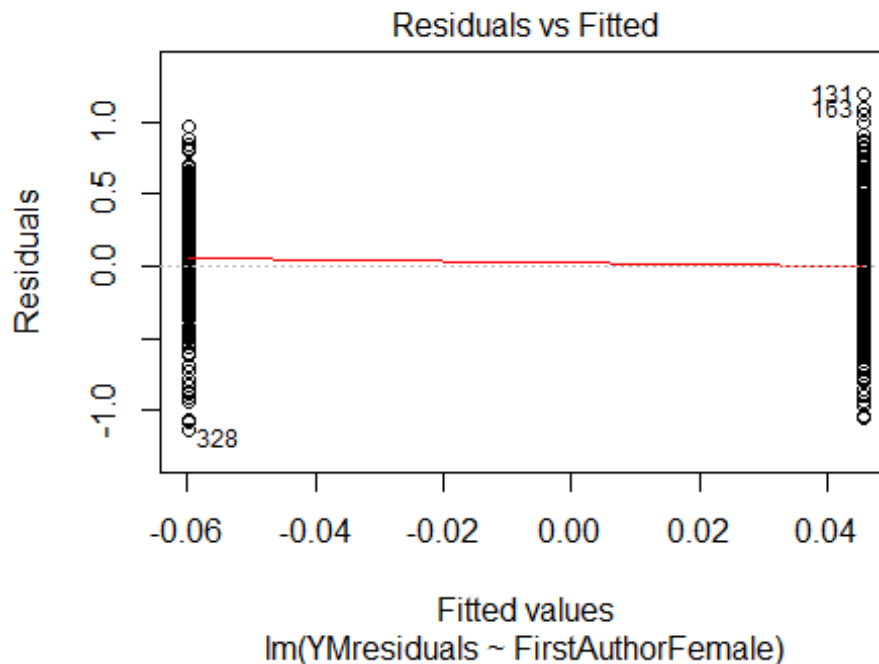


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 5.89858397559995"
## [1] "Male first author team size 2018 geometric mean: 5.07278674892196"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



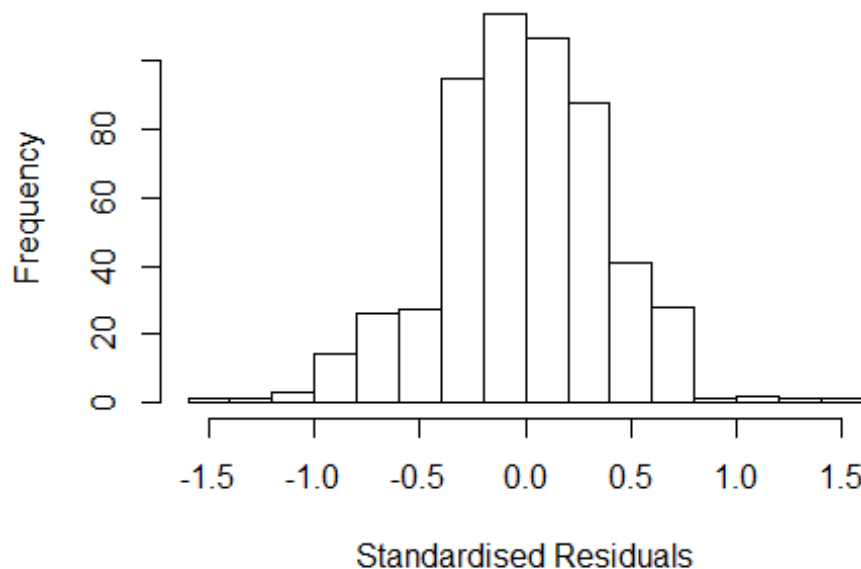
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.55321168438643"
## [1] "Male last author team size 2018 geometric mean: 5.13346314609778"

## 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 = 84, 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 1.309 1      1.144
## LastAuthorFemale  1.672 1      1.293
## UniqueAuthors    2.841 4      1.139
## Year              2.546 16     1.030
```

Residuals from first and last author and team size



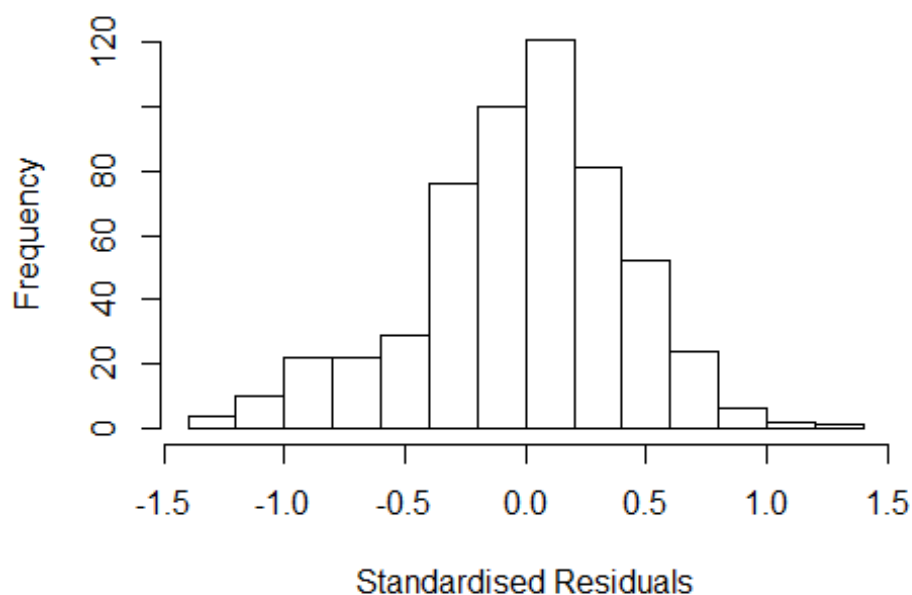
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40363 -0.23553 -0.00345  0.26123  1.47992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.25e-01   2.19e-01   1.94  0.05308 .
## FirstAuthorFemale1 -6.89e-02   3.60e-02  -1.91  0.05635 .
## LastAuthorFemale1 -1.24e-01   4.90e-02  -2.52  0.01190 *
## UniqueAuthors2     5.27e-01   1.39e-01   3.78  0.00017 ***
## UniqueAuthors3     6.17e-01   1.17e-01   5.27  2.0e-07 ***
## UniqueAuthors4     6.34e-01   1.13e-01   5.60  3.5e-08 ***
## UniqueAuthors5     8.13e-01   1.01e-01   8.06  5.0e-15 ***
## Year1997           2.87e-02   2.07e-01   0.14  0.89015
## Year1998          -1.78e-01   2.03e-01  -0.87  0.38304
## Year1999          -1.36e-02   2.10e-01  -0.06  0.94843
```

```

## Year2000          3.19e-02    2.05e-01    0.16  0.87626
## Year2001          1.14e-01    2.17e-01    0.53  0.59977
## Year2002          1.34e-01    2.06e-01    0.65  0.51604
## Year2003         -2.20e-02    2.01e-01   -0.11  0.91284
## Year2004          2.17e-02    2.05e-01    0.11  0.91607
## Year2005         -5.24e-02    2.00e-01   -0.26  0.79315
## Year2006         -3.12e-05    1.97e-01    0.00  0.99987
## Year2007          1.23e-01    2.00e-01    0.61  0.53976
## Year2008          2.34e-01    2.09e-01    1.12  0.26314
## Year2009          9.16e-02    2.06e-01    0.45  0.65618
## Year2010         -1.15e-02    1.99e-01   -0.06  0.95403
## Year2011          7.48e-03    2.02e-01    0.04  0.97052
## Year2012         -1.47e-01    1.99e-01   -0.74  0.45979
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.31,   Adjusted R-squared:  0.281
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 505 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0321 0.8640 0.9460 0.8870 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          1.82e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.243 1          1.115
## LastAuthorFemale 1.380 1          1.175
## Year          1.568 16          1.014

```

Residuals from first and last author



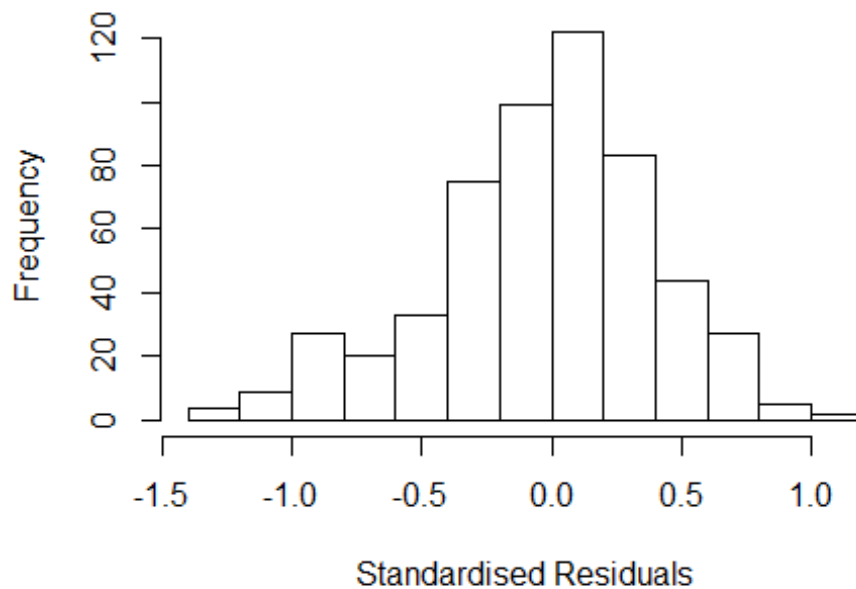
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3376 -0.2559 0.0211 0.2606 1.3505
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13093 0.17328 6.53 1.6e-10 ***
## FirstAuthorFemale1 -0.07248 0.03957 -1.83 0.0676 .
## LastAuthorFemale1 -0.19202 0.05835 -3.29 0.0011 **
## Year1997 0.07304 0.19375 0.38 0.7063
## Year1998 -0.15845 0.19443 -0.81 0.4155
## Year1999 0.00675 0.20211 0.03 0.9734
## Year2000 0.08573 0.18869 0.45 0.6497
## Year2001 0.21100 0.19508 1.08 0.2799
## Year2002 0.19897 0.18524 1.07 0.2833
## Year2003 0.02957 0.18371 0.16 0.8722
## Year2004 0.06908 0.18818 0.37 0.7137
## Year2005 -0.01617 0.17994 -0.09 0.9284
```

```

## Year2006          0.02929    0.17928    0.16    0.8703
## Year2007          0.16239    0.18195    0.89    0.3725
## Year2008          0.27919    0.18996    1.47    0.1422
## Year2009          0.14177    0.18784    0.75    0.4507
## Year2010         -0.04397    0.19289   -0.23    0.8198
## Year2011         -0.05764    0.19554   -0.29    0.7683
## Year2012         -0.11505    0.18320   -0.63    0.5303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.115, Adjusted R-squared:  0.0854
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 506 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.186  0.853   0.949   0.886   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      1.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.33 1         1.153
## Year              1.33 16         1.009

```

Residuals from first author



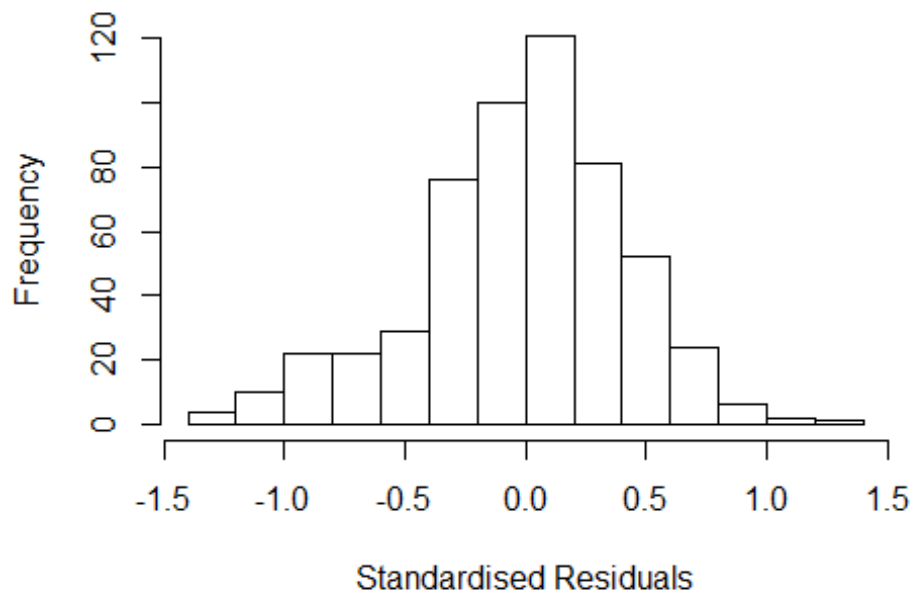
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2703 -0.2644 0.0147 0.2582 1.1542
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0888 0.1775 6.13 1.7e-09 ***
## FirstAuthorFemale1 -0.0793 0.0427 -1.86 0.064 .
## Year1997 0.1009 0.1960 0.51 0.607
## Year1998 -0.1064 0.1981 -0.54 0.592
## Year1999 -0.0186 0.2139 -0.09 0.931
## Year2000 0.1183 0.1914 0.62 0.537
## Year2001 0.1889 0.1967 0.96 0.337
## Year2002 0.1914 0.1907 1.00 0.316
## Year2003 0.0350 0.1863 0.19 0.851
## Year2004 0.0840 0.1926 0.44 0.663
## Year2005 -0.0123 0.1835 -0.07 0.947
## Year2006 0.0381 0.1831 0.21 0.835
```

```

## Year2007          0.1878      0.1866      1.01      0.315
## Year2008          0.2608      0.1945      1.34      0.181
## Year2009          0.1510      0.1922      0.79      0.432
## Year2010         -0.0262      0.1995     -0.13      0.895
## Year2011         -0.0475      0.2060     -0.23      0.818
## Year2012         -0.1191      0.1888     -0.63      0.528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0732, Adjusted R-squared:  0.0436
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.245  0.856  0.951  0.885  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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.383 1      1.176
## Year            1.383 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
## -1.3805 -0.2519 0.0103 0.2689 1.3655
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1153 0.1735 6.43 2.9e-10 ***
## LastAuthorFemale1 -0.1975 0.0608 -3.25 0.0012 **
## Year1997 0.0527 0.1952 0.27 0.7872
## Year1998 -0.1524 0.1954 -0.78 0.4357
## Year1999 -0.0199 0.2030 -0.10 0.9220
## Year2000 0.0715 0.1892 0.38 0.7058
## Year2001 0.1836 0.1940 0.95 0.3443
## Year2002 0.1916 0.1865 1.03 0.3046
## Year2003 0.0252 0.1842 0.14 0.8911
## Year2004 0.0499 0.1886 0.26 0.7916
## Year2005 -0.0284 0.1803 -0.16 0.8750
## Year2006 0.0191 0.1796 0.11 0.9155
```

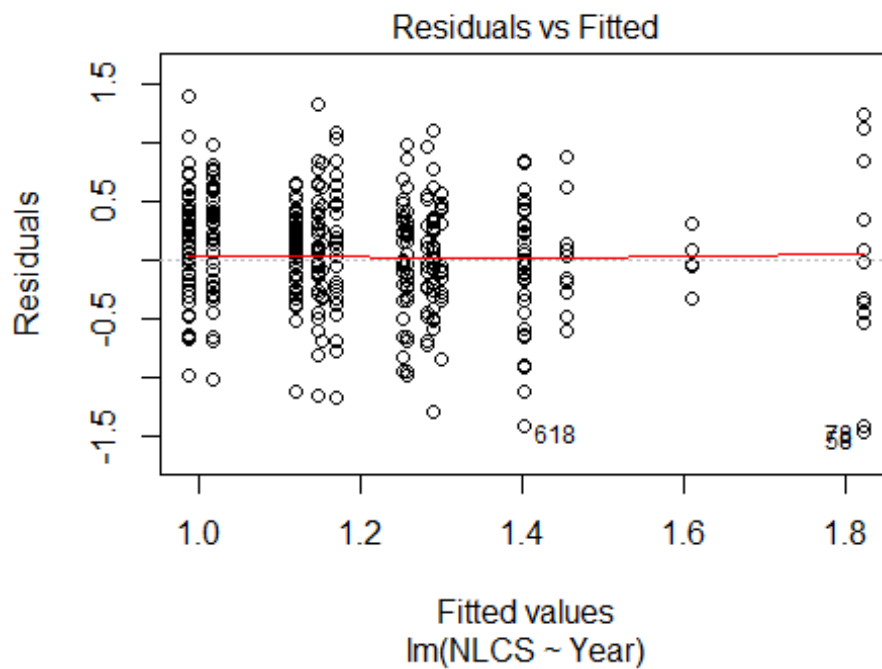


```

## Year2007          0.1356      0.1824      0.74      0.4576
## Year2008          0.2652      0.1899      1.40      0.1631
## Year2009          0.1368      0.1893      0.72      0.4701
## Year2010         -0.0549      0.1952     -0.28      0.7787
## Year2011         -0.0759      0.1972     -0.38      0.7005
## Year2012         -0.1346      0.1841     -0.73      0.4651
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0794
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 505 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.165  0.854  0.950  0.885  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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 550"
## [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
##   38   47   36   33   42   38   38   25   26   39   47   64   59   50   53
## 2011 2012
##  117   66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   13   12   14   11    5   14   12   16   21   30   41   40   34   35
## 2011 2012

```

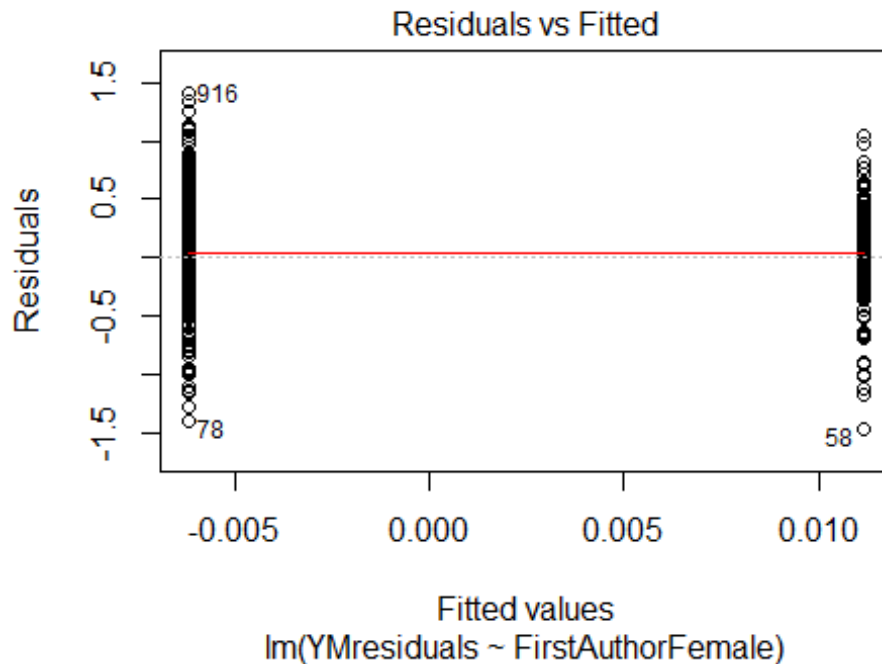
```
## 75 52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 12 11 12 10 4 12 10 16 20 28 38 34 28 33
## 2011 2012
## 65 45
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.2, df = 1, p-value = 0.004

## [1] "Female first author team size 2018 geometric mean: 4.59153832008575"
## [1] "Male first author team size 2018 geometric mean: 5.24849323179734"

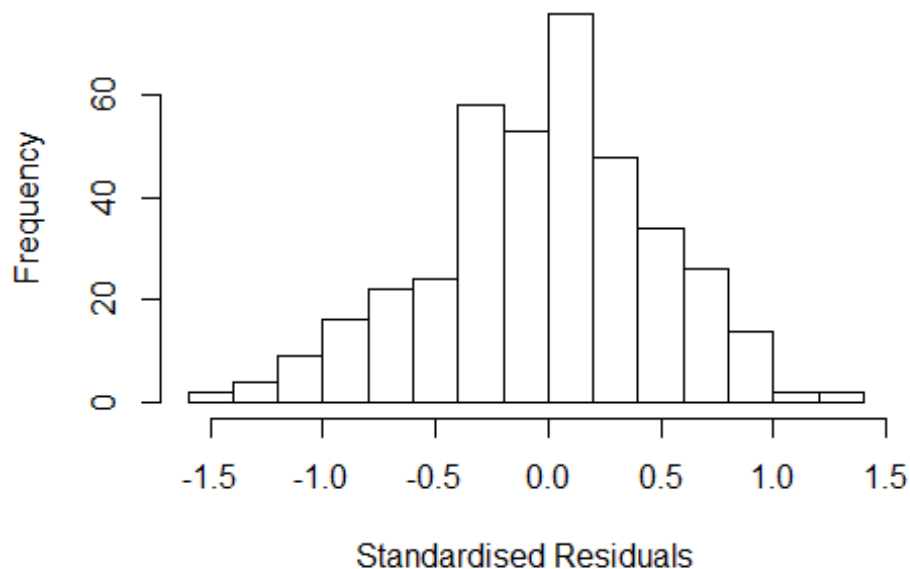
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 490, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.03401423243322"
## [1] "Male last author team size 2018 geometric mean: 4.9940425725502"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 460, 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	1.342	1	1.158
LastAuthorFemale	1.117	1	1.057
UniqueAuthors	2.709	4	1.133
Year	3.319	16	1.038

Residuals from first and last author and team size



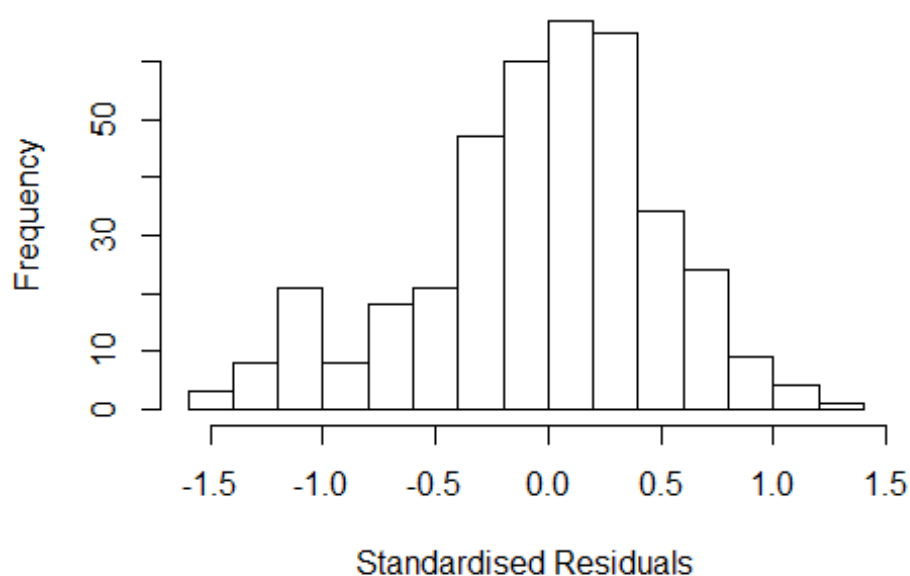
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5187 -0.3152 0.0223 0.2899 1.2460
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.79238 0.20297 3.90 0.00011 ***
## FirstAuthorFemale1 -0.01075 0.05544 -0.19 0.84632
## LastAuthorFemale1 0.01094 0.05354 0.20 0.83815
## UniqueAuthors2 0.21243 0.15526 1.37 0.17209
## UniqueAuthors3 0.43056 0.15707 2.74 0.00642 **
## UniqueAuthors4 0.45799 0.13995 3.27 0.00117 **
## UniqueAuthors5 0.56174 0.13462 4.17 3.8e-05 ***
## Year1997 0.67349 0.45012 1.50 0.13544
## Year1998 0.29858 0.22763 1.31 0.19043
## Year1999 -0.00858 0.28756 -0.03 0.97622
```

```

## Year2000      -0.02008    0.25254   -0.08   0.93668
## Year2001      0.57033    0.21866    2.61   0.00947 **
## Year2002      0.04795    0.22514    0.21   0.83145
## Year2003     -0.05318    0.18290   -0.29   0.77140
## Year2004      0.06675    0.19348    0.35   0.73027
## Year2005      0.00634    0.18610    0.03   0.97285
## Year2006      0.06497    0.18162    0.36   0.72075
## Year2007     -0.09420    0.17396   -0.54   0.58851
## Year2008      0.28782    0.18777    1.53   0.12617
## Year2009     -0.03550    0.19349   -0.18   0.85451
## Year2010      0.05915    0.17617    0.34   0.73726
## Year2011     -0.15580    0.17357   -0.90   0.36998
## Year2012     -0.12101    0.18914   -0.64   0.52271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.209, Adjusted R-squared:  0.161
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 362 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.843   0.954   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      2.56e-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 1.329 1      1.153
## LastAuthorFemale  1.092 1      1.045
## Year              1.450 16      1.012

```

Residuals from first and last author



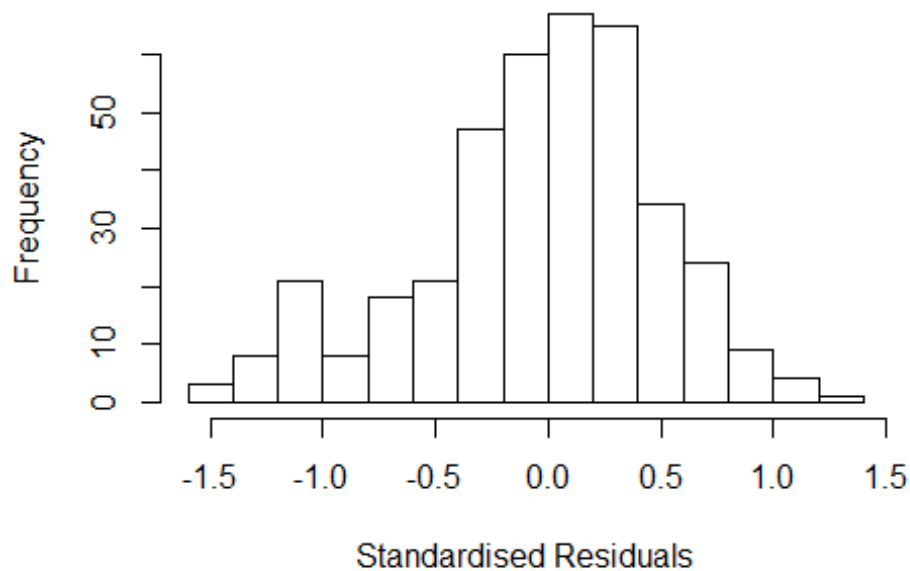
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5229 -0.3027  0.0251  0.3146  1.2473
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.217747   0.212806   5.72 2.2e-08 ***
## FirstAuthorFemale1 -0.006376   0.057378  -0.11  0.912
## LastAuthorFemale1 -0.000487   0.053122  -0.01  0.993
## Year1997         0.678559   0.500651   1.36  0.176
## Year1998         0.273315   0.241477   1.13  0.258
## Year1999         0.048595   0.304650   0.16  0.873
## Year2000        -0.000777   0.266404   0.00  0.998
## Year2001         0.408177   0.243081   1.68  0.094 .
## Year2002         0.069525   0.247232   0.28  0.779
## Year2003        -0.079426   0.220923  -0.36  0.719
## Year2004         0.056816   0.243415   0.23  0.816
## Year2005         0.033626   0.234225   0.14  0.886
```

```

## Year2006          0.051241    0.228349    0.22    0.823
## Year2007          -0.062916    0.222894   -0.28    0.778
## Year2008          0.271034    0.233559    1.16    0.247
## Year2009          0.011946    0.242962    0.05    0.961
## Year2010          0.113484    0.224899    0.50    0.614
## Year2011          -0.199592    0.222560   -0.90    0.370
## Year2012          -0.139273    0.237534   -0.59    0.558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0681
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 358 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.280  0.837  0.953  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      2.56e-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 1.321 1      1.149
## Year              1.321 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
## -1.5219 -0.3027  0.0256  0.3149  1.2478
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.216960   0.212583   5.72 2.1e-08 ***
## FirstAuthorFemale1 -0.006375   0.058858  -0.11  0.914
## Year1997         0.678316   0.497127   1.36  0.173
## Year1998         0.274156   0.241219   1.14  0.256
## Year1999         0.049220   0.304122   0.16  0.872
## Year2000        -0.000147   0.265556   0.00  1.000
## Year2001         0.408956   0.242962   1.68  0.093 .
## Year2002         0.070127   0.246470   0.28  0.776
## Year2003        -0.078840   0.219991  -0.36  0.720
## Year2004         0.057540   0.242822   0.24  0.813
## Year2005         0.034091   0.233423   0.15  0.884
## Year2006         0.051813   0.227693   0.23  0.820
```

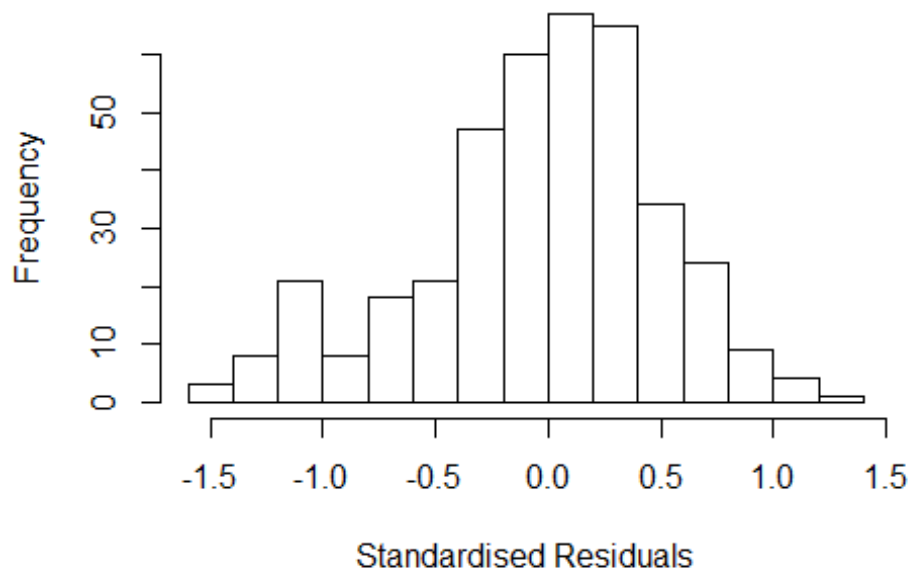


```

## Year2007      -0.062683    0.221790   -0.28    0.778
## Year2008      0.271236    0.232668    1.17    0.244
## Year2009      0.012198    0.242344    0.05    0.960
## Year2010      0.113877    0.223710    0.51    0.611
## Year2011     -0.199275    0.221828   -0.90    0.370
## Year2012     -0.139149    0.236652   -0.59    0.557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.475
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0704
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 358 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.839  0.953  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      2.56e-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 1.091 1      1.045
## Year              1.091 16      1.003

```

Residuals from last author



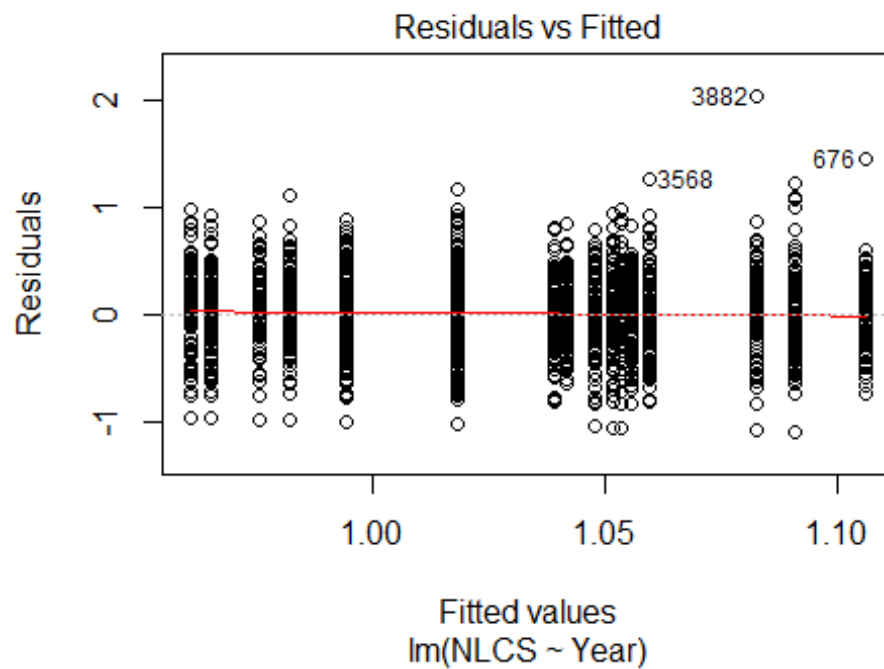
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5317 -0.3071 0.0272 0.3152 1.2501
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21761 0.21233 5.73 2e-08 ***
## LastAuthorFemale1 -0.00244 0.05526 -0.04 0.965
## Year1997 0.68114 0.49848 1.37 0.173
## Year1998 0.27106 0.24153 1.12 0.262
## Year1999 0.04694 0.30514 0.15 0.878
## Year2000 -0.00146 0.26683 -0.01 0.996
## Year2001 0.40663 0.24347 1.67 0.096 .
## Year2002 0.06830 0.24753 0.28 0.783
## Year2003 -0.08170 0.22124 -0.37 0.712
## Year2004 0.05485 0.24393 0.22 0.822
## Year2005 0.03177 0.23441 0.14 0.892
## Year2006 0.05038 0.22867 0.22 0.826
```

```

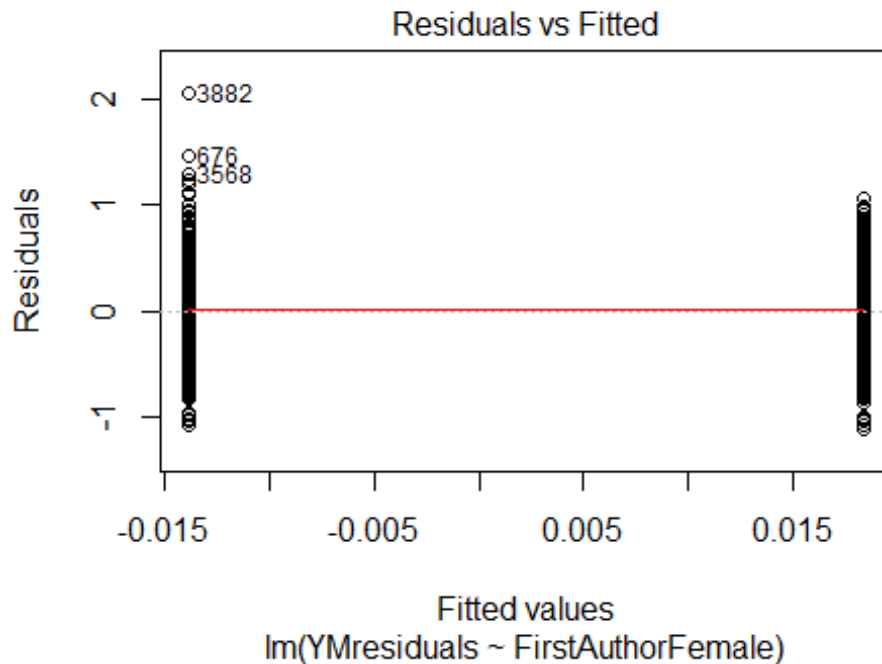
## Year2007          -0.06481    0.22312   -0.29    0.772
## Year2008           0.26846    0.23331    1.15    0.251
## Year2009           0.00930    0.24112    0.04    0.969
## Year2010           0.11297    0.22515    0.50    0.616
## Year2011          -0.20085    0.22322   -0.90    0.369
## Year2012          -0.14177    0.23620   -0.60    0.549
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0708
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 359 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.272  0.837   0.953   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      2.56e-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: 390"
## [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
##   201  197  220  211  190  216  171  180  208  223  253  213  219  181  168
## 2011 2012
##   166  207
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    96   95  112  111   94   94  110  119  124  137  165  139  155  121  115
## 2011 2012

```

```
## 116 143
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 85 103 96 87 84 100 103 118 117 141 116 136 113 104
## 2011 2012
## 109 135
## [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
```

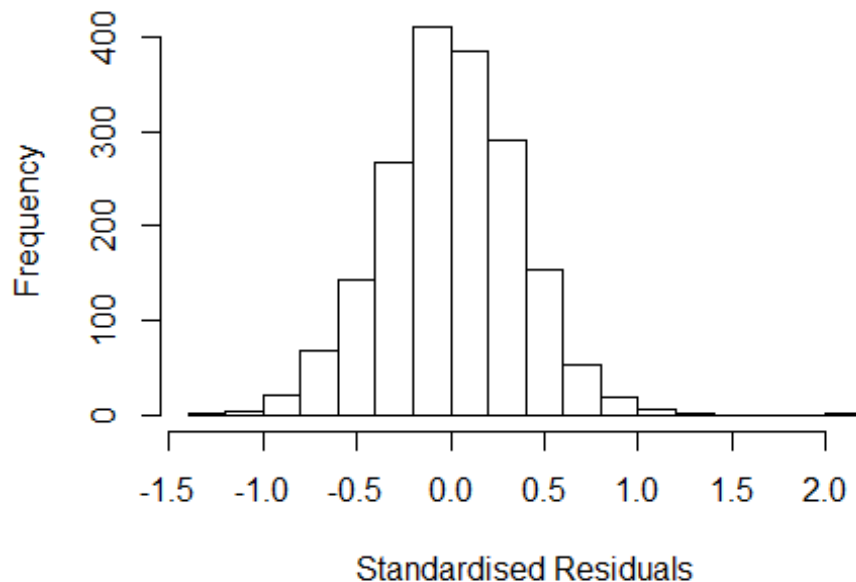


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.44716604590564"
## [1] "Male first author team size 2018 geometric mean: 3.86888991941501"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.28168391275473"
## [1] "Male last author team size 2018 geometric mean: 4.0552380208832"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, 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 1.080 1      1.039
## LastAuthorFemale  1.063 1      1.031
## UniqueAuthors    1.282 4      1.032
## Year              1.354 16     1.010
```

Residuals from first and last author and team size



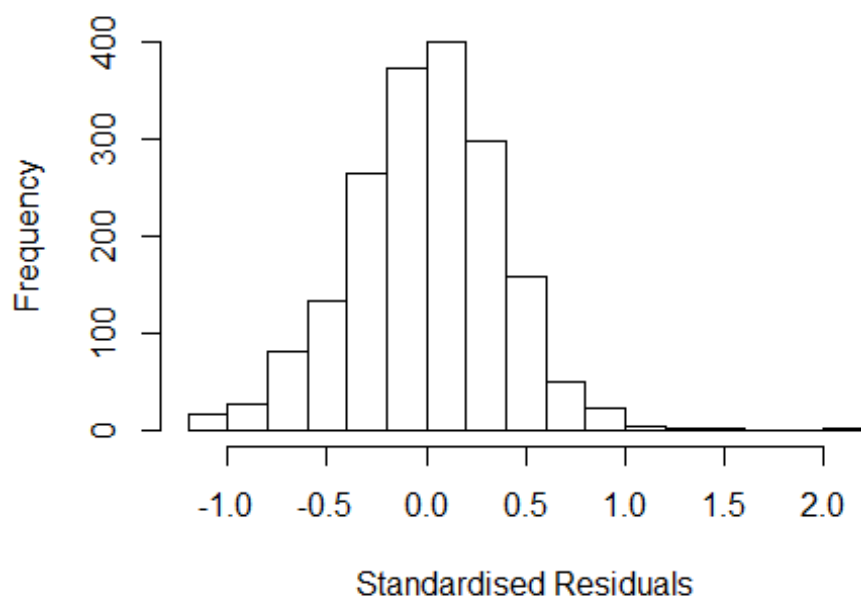
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22316 -0.23250 -0.00191  0.24448  2.06205
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.81453    0.06159   13.23 < 2e-16 ***
## FirstAuthorFemale1 0.00752    0.01745    0.43  0.6666
## LastAuthorFemale1 -0.02037    0.01977   -1.03  0.3029
## UniqueAuthors2    0.24863    0.04850    5.13 3.3e-07 ***
## UniqueAuthors3    0.26526    0.04781    5.55 3.3e-08 ***
## UniqueAuthors4    0.32351    0.04947    6.54 8.0e-11 ***
## UniqueAuthors5    0.40645    0.04821    8.43 < 2e-16 ***
## Year1997          0.02254    0.06327    0.36  0.7216
## Year1998          0.03131    0.05389    0.58  0.5613
## Year1999         -0.05048    0.05444   -0.93  0.3539
```

```

## Year2000      -0.05815      0.05846      -0.99      0.3201
## Year2001      -0.11181      0.05996      -1.86      0.0624 .
## Year2002      -0.11196      0.05828      -1.92      0.0549 .
## Year2003      -0.14492      0.05551      -2.61      0.0091 **
## Year2004      -0.10102      0.05404      -1.87      0.0617 .
## Year2005      -0.05911      0.05232      -1.13      0.2587
## Year2006      -0.13589      0.05323      -2.55      0.0108 *
## Year2007      -0.06103      0.05492      -1.11      0.2666
## Year2008      -0.09694      0.05508      -1.76      0.0786 .
## Year2009      -0.07610      0.05309      -1.43      0.1519
## Year2010      -0.06518      0.05789      -1.13      0.2604
## Year2011      -0.00522      0.05504      -0.09      0.9244
## Year2012      -0.08906      0.05744      -1.55      0.1212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.352
## Multiple R-squared:  0.0919, Adjusted R-squared:  0.0808
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 1671 is an outlier with |weight| = 0 ( < 5.5e-05);
## 153 weights are ~= 1. The remaining 1676 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.136  0.869  0.950  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           5.46e-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 1.099 1 1.048
## LastAuthorFemale 1.038 1 1.019
## Year 1.107 16 1.003

```

Residuals from first and last author



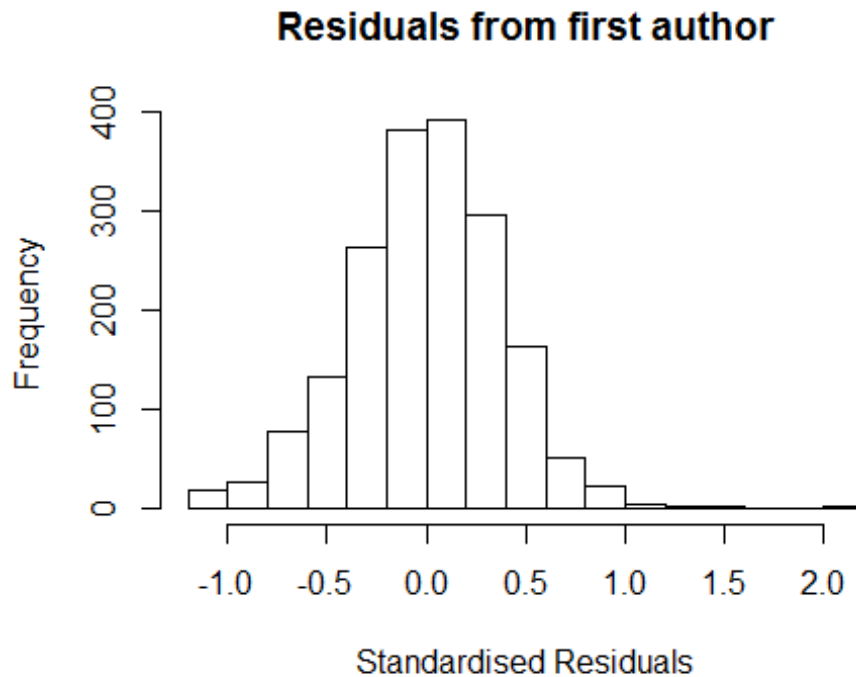
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10152 -0.23785 0.00849 0.24276 2.01848
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.074465 0.043627 24.63 <2e-16 ***
## FirstAuthorFemale1 0.027785 0.018074 1.54 0.124
## LastAuthorFemale1 -0.024603 0.019953 -1.23 0.218
## Year1997 0.000304 0.063916 0.00 0.996
## Year1998 0.017112 0.052778 0.32 0.746
## Year1999 -0.060475 0.053124 -1.14 0.255
## Year2000 -0.048500 0.057078 -0.85 0.396
## Year2001 -0.097716 0.061860 -1.58 0.114
## Year2002 -0.106203 0.059700 -1.78 0.075 .
## Year2003 -0.143143 0.056909 -2.52 0.012 *
## Year2004 -0.077995 0.055442 -1.41 0.160
## Year2005 -0.030278 0.052208 -0.58 0.562
```



```

## Year2006          -0.107582    0.054412   -1.98    0.048 *
## Year2007          -0.027425    0.054707   -0.50    0.616
## Year2008          -0.054105    0.055546   -0.97    0.330
## Year2009          -0.036758    0.051912   -0.71    0.479
## Year2010          -0.025787    0.059445   -0.43    0.664
## Year2011           0.027058    0.056316    0.48    0.631
## Year2012          -0.046084    0.057959   -0.80    0.427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.00657
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 1671 is an outlier with |weight| = 0 ( < 5.5e-05);
## 160 weights are ~= 1. The remaining 1669 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0545 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      5.46e-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 1.082 1          1.040
## Year              1.082 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
## -1.09592 -0.23832  0.00825  0.24343  2.02408
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07e+00  4.35e-02  24.64  <2e-16 ***
## FirstAuthorFemale1 2.41e-02  1.80e-02   1.34   0.180
## Year1997        -2.65e-05  6.39e-02   0.00   1.000
## Year1998         1.62e-02  5.29e-02   0.31   0.760
## Year1999        -6.16e-02  5.31e-02  -1.16   0.246
## Year2000        -4.72e-02  5.71e-02  -0.83   0.408
## Year2001        -9.59e-02  6.17e-02  -1.55   0.120
## Year2002       -1.07e-01  5.97e-02  -1.79   0.074 .
## Year2003       -1.44e-01  5.68e-02  -2.54   0.011 *
## Year2004       -7.82e-02  5.53e-02  -1.42   0.157
## Year2005       -3.14e-02  5.21e-02  -0.60   0.546
## Year2006       -1.10e-01  5.43e-02  -2.02   0.044 *
```

```

## Year2007      -2.91e-02   5.45e-02   -0.53   0.593
## Year2008      -5.59e-02   5.53e-02   -1.01   0.312
## Year2009      -3.59e-02   5.18e-02   -0.69   0.488
## Year2010      -2.74e-02   5.93e-02   -0.46   0.644
## Year2011       2.50e-02   5.62e-02    0.45   0.656
## Year2012      -4.72e-02   5.78e-02   -0.82   0.414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.00638
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1671 is an outlier with |weight| = 0 ( < 5.5e-05);
## 153 weights are ~= 1. The remaining 1676 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0534 0.8690 0.9510 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      5.46e-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 1.025 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

```

```

## -1.11336 -0.23518 0.00976 0.24066 2.00664
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.08460    0.04255   25.49  <2e-16 ***
## LastAuthorFemale1 -0.01902    0.01983   -0.96   0.338
## Year1997          0.00241    0.06380    0.04   0.970
## Year1998          0.01439    0.05234    0.27   0.783
## Year1999         -0.06413    0.05277   -1.22   0.224
## Year2000         -0.04442    0.05670   -0.78   0.433
## Year2001         -0.09262    0.06094   -1.52   0.129
## Year2002         -0.10640    0.05932   -1.79   0.073 .
## Year2003         -0.14170    0.05673   -2.50   0.013 *
## Year2004         -0.07753    0.05512   -1.41   0.160
## Year2005         -0.03095    0.05195   -0.60   0.551
## Year2006         -0.10607    0.05414   -1.96   0.050 .
## Year2007         -0.02545    0.05446   -0.47   0.640
## Year2008         -0.05422    0.05511   -0.98   0.325
## Year2009         -0.03525    0.05157   -0.68   0.494
## Year2010         -0.02273    0.05917   -0.38   0.701
## Year2011          0.02876    0.05585    0.51   0.607
## Year2012         -0.04326    0.05756   -0.75   0.452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00581
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1671 is an outlier with |weight| = 0 ( < 5.5e-05);
## 150 weights are ~ = 1. The remaining 1679 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0576 0.8700 0.9510 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           5.46e-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)

```

```

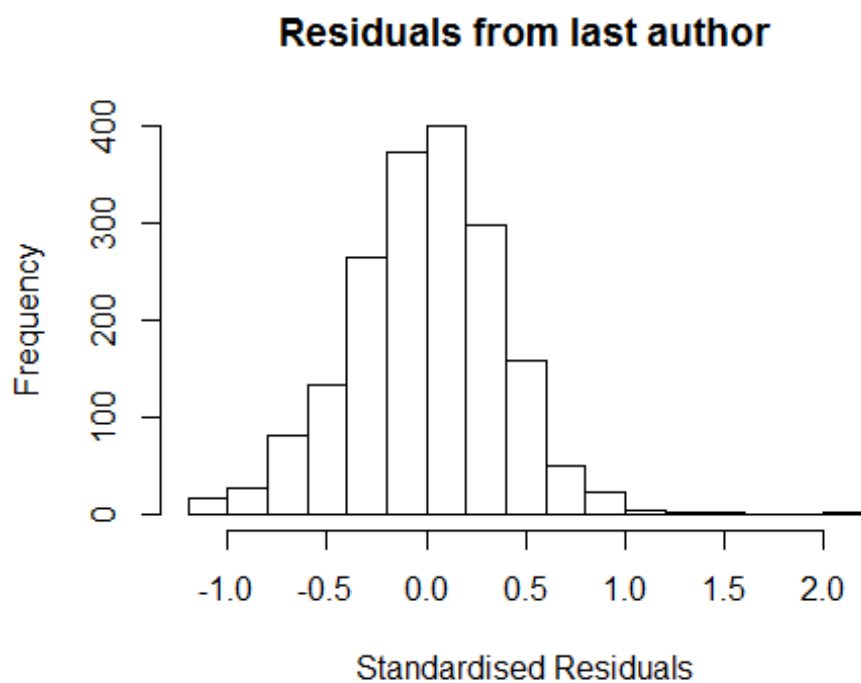
## [1] "Sample size for the above analysis: 1830"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2801"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    3    7    8    8    4    5    9   10   15   11   12   10   24   27
## 2011 2012
##   19   23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    1    5    7    2    0    5    8    9    8   11   10    8   22   22
## 2011 2012
##   13   18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    1    5    6    2    0    5    7    7    6   10    9    7   19   20
## 2011 2012
##   11   18
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5.65763545978916"
## [1] "Male first author team size 2018 geometric mean: 5.23673426566977"

## 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 = 76, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.51915072633213"
## [1] "Male last author team size 2018 geometric mean: 5.55532509258829"

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

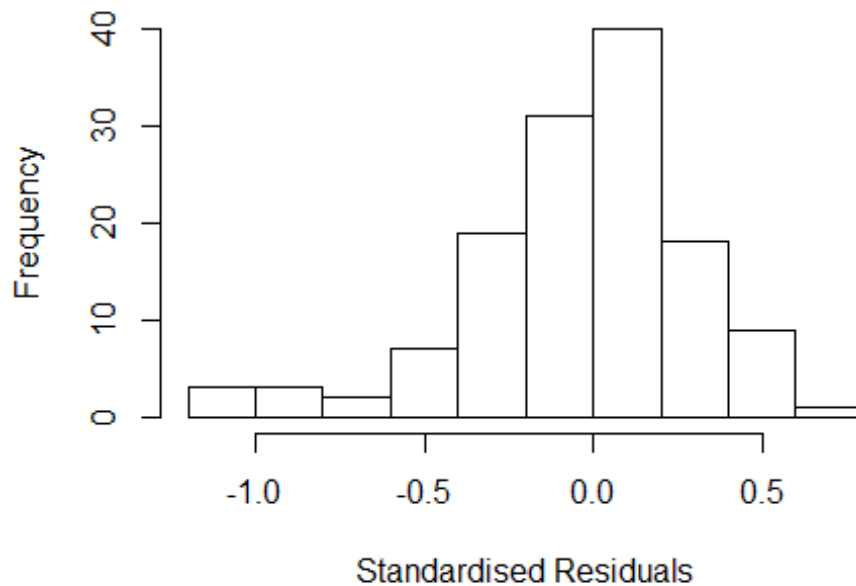
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 84, 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	2.778	1	1.667
## LastAuthorFemale	3.621	1	1.903
## UniqueAuthors	65.142	4	1.686
## Year	151.045	14	1.196

Residuals from first and last author and team size



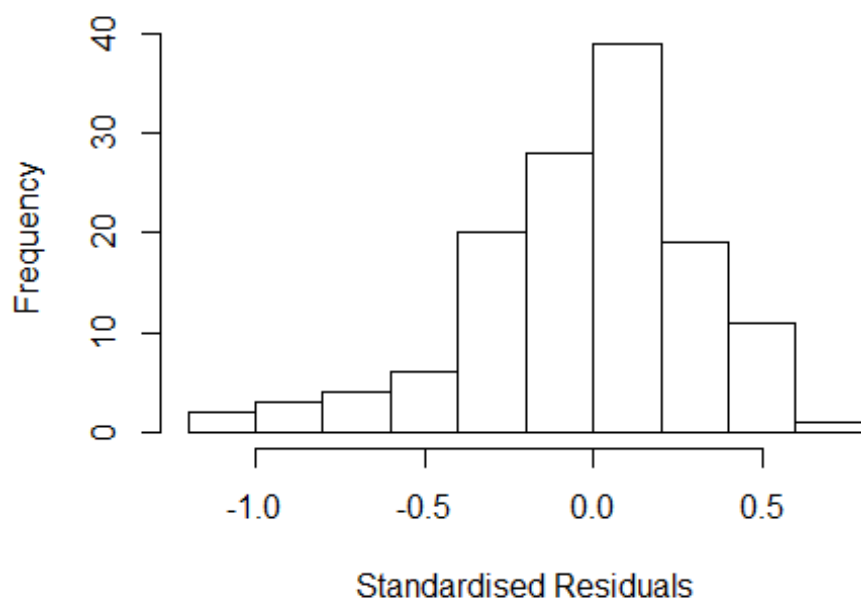
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.03875 -0.20580 0.00513 0.16688 0.67583
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3454 0.2806 4.79 5.1e-06 ***
## FirstAuthorFemale1 0.0502 0.0677 0.74 0.4598
## LastAuthorFemale1 -0.0196 0.0611 -0.32 0.7492
## UniqueAuthors2 -0.0876 0.2525 -0.35 0.7293
## UniqueAuthors3 0.0349 0.2401 0.15 0.8846
## UniqueAuthors4 -0.0364 0.2703 -0.13 0.8932
## UniqueAuthors5 0.2115 0.2323 0.91 0.3645
## Year1998 -0.2148 0.2714 -0.79 0.4304
## Year1999 -0.3067 0.2085 -1.47 0.1442
## Year2000 -0.2834 0.3738 -0.76 0.4499
```

```

## Year2002          -0.3066      0.1306    -2.35    0.0206 *
## Year2003          -0.2979      0.1916    -1.55    0.1228
## Year2004          -0.3435      0.0973    -3.53    0.0006 ***
## Year2005          -0.4632      0.1627    -2.85    0.0053 **
## Year2006          -0.2489      0.1360    -1.83    0.0699 .
## Year2007          -0.2999      0.1315    -2.28    0.0245 *
## Year2008          -0.2566      0.1478    -1.74    0.0853 .
## Year2009          -0.2113      0.0817    -2.59    0.0110 *
## Year2010          -0.3418      0.1312    -2.61    0.0104 *
## Year2011          -0.3719      0.1124    -3.31    0.0013 **
## Year2012          -0.3593      0.0889    -4.04    9.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.285
## Multiple R-squared:  0.182, Adjusted R-squared:  0.0364
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.157  0.862  0.949  0.878  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.52e-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 1.967 1          1.402
## LastAuthorFemale  3.010 1          1.735
## Year              5.341 14          1.062

```


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
## -1.1977 -0.2111  0.0238  0.1686  0.7486
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2526    0.0686   18.26  <2e-16 ***
## FirstAuthorFemale1  0.0554    0.0686    0.81  0.4211
## LastAuthorFemale1 -0.0536    0.0602   -0.89  0.3752
## Year1998          -0.1031    0.2104   -0.49  0.6250
## Year1999          -0.2767    0.2009   -1.38  0.1710
## Year2000          -0.2212    0.3826   -0.58  0.5642
## Year2002          -0.2061    0.1222   -1.69  0.0945 .
## Year2003          -0.1873    0.1831   -1.02  0.3083
## Year2004          -0.1199    0.0737   -1.63  0.1065
## Year2005          -0.2462    0.1918   -1.28  0.2019
## Year2006          -0.1027    0.1217   -0.84  0.4008
## Year2007          -0.0455    0.0839   -0.54  0.5889
```

```

## Year2008          -0.0989      0.1732   -0.57    0.5691
## Year2009          -0.0324      0.0654   -0.49    0.6216
## Year2010          -0.1103      0.0869   -1.27    0.2071
## Year2011          -0.2203      0.0820   -2.69    0.0082 **
## Year2012          -0.2763      0.0851   -3.25    0.0015 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared:  0.0846, Adjusted R-squared:  -0.0417
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.144  0.882  0.959  0.897  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      7.52e-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 1.879 1          1.371
## Year              1.879 14          1.023

## [1] "List of 0 outliers 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
## -1.1860 -0.2090  0.0126  0.1797  0.7551
##

```

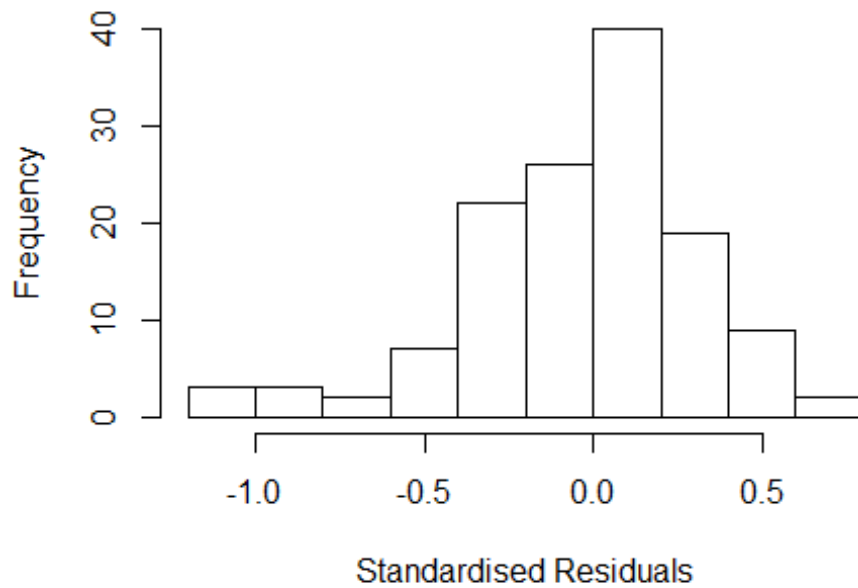
```

## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2579    0.0689   18.25 < 2e-16 ***
## FirstAuthorFemale1 0.0501    0.0689    0.73 0.46871
## Year1998         -0.1420    0.1989   -0.71 0.47660
## Year1999         -0.3074    0.2064   -1.49 0.13901
## Year2000         -0.2774    0.3749   -0.74 0.46077
## Year2002         -0.2218    0.1203   -1.84 0.06775 .
## Year2003         -0.2040    0.1826   -1.12 0.26609
## Year2004         -0.1524    0.0616   -2.47 0.01477 *
## Year2005         -0.2779    0.1926   -1.44 0.15166
## Year2006         -0.1384    0.0970   -1.43 0.15612
## Year2007         -0.0767    0.0782   -0.98 0.32849
## Year2008         -0.1255    0.1756   -0.71 0.47645
## Year2009         -0.0516    0.0637   -0.81 0.41983
## Year2010         -0.1220    0.0869   -1.40 0.16309
## Year2011         -0.2506    0.0695   -3.61 0.00046 ***
## Year2012         -0.3027    0.0763   -3.97 0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.0799, Adjusted R-squared:  -0.0381
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.143  0.875  0.960  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      7.52e-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"

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

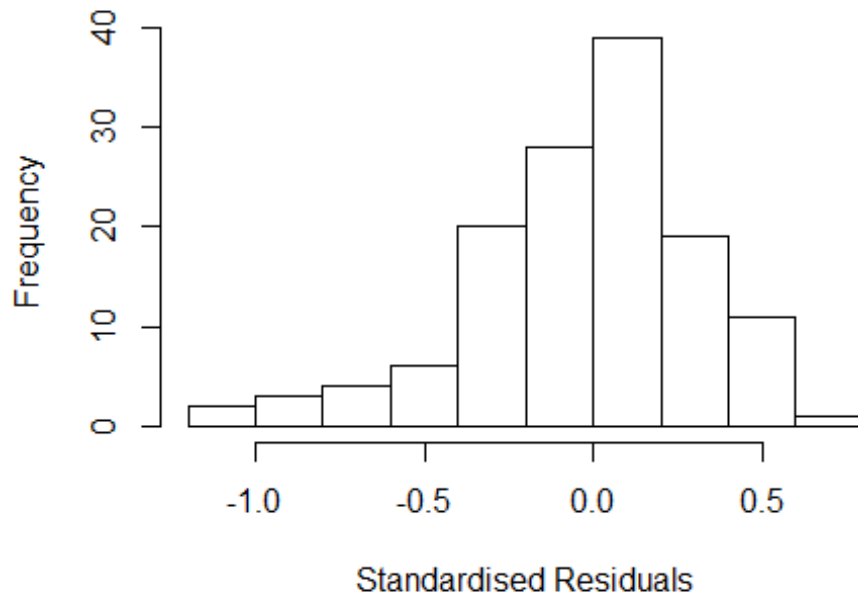
```

Residuals from first author



##	GVIF	Df	$GVIF^{(1/(2*Df))}$
## LastAuthorFemale	2.724	1	1.651
## Year	2.724	14	1.036

Residuals from last author

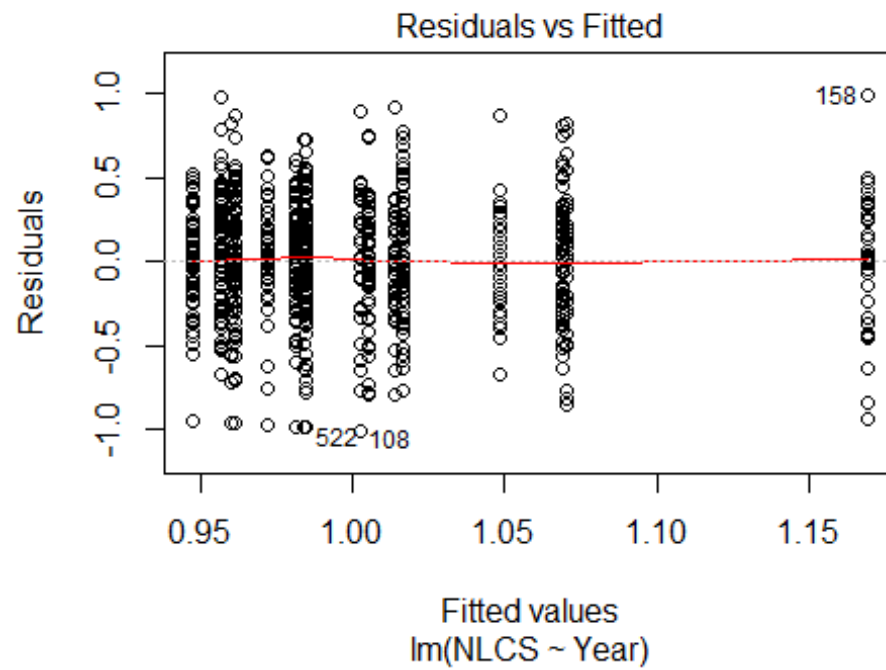


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1672 -0.2207 0.0237 0.1825 0.7238
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3080 0.0000 Inf < 2e-16 ***
## LastAuthorFemale1 -0.0493 0.0594 -0.83 0.40822
## Year1998 -0.1079 0.2227 -0.48 0.62884
## Year1999 -0.3152 0.2053 -1.54 0.12729
## Year2000 -0.2532 0.3350 -0.76 0.45125
## Year2002 -0.2363 0.1191 -1.98 0.04966 *
## Year2003 -0.2436 0.1684 -1.45 0.15077
## Year2004 -0.1380 0.0699 -1.98 0.05053 .
## Year2005 -0.2653 0.1762 -1.51 0.13483
## Year2006 -0.1183 0.1136 -1.04 0.30005
## Year2007 -0.0725 0.0682 -1.06 0.29015
## Year2008 -0.1047 0.1663 -0.63 0.52998
## Year2009 -0.0615 0.0598 -1.03 0.30582
## Year2010 -0.1408 0.0904 -1.56 0.12217
## Year2011 -0.2348 0.0861 -2.73 0.00733 **
## Year2012 -0.2978 0.0839 -3.55 0.00055 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared: 0.079, Adjusted R-squared: -0.0391
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~ 1. The remaining 122 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.167 0.887 0.959 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 7.52e-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
```

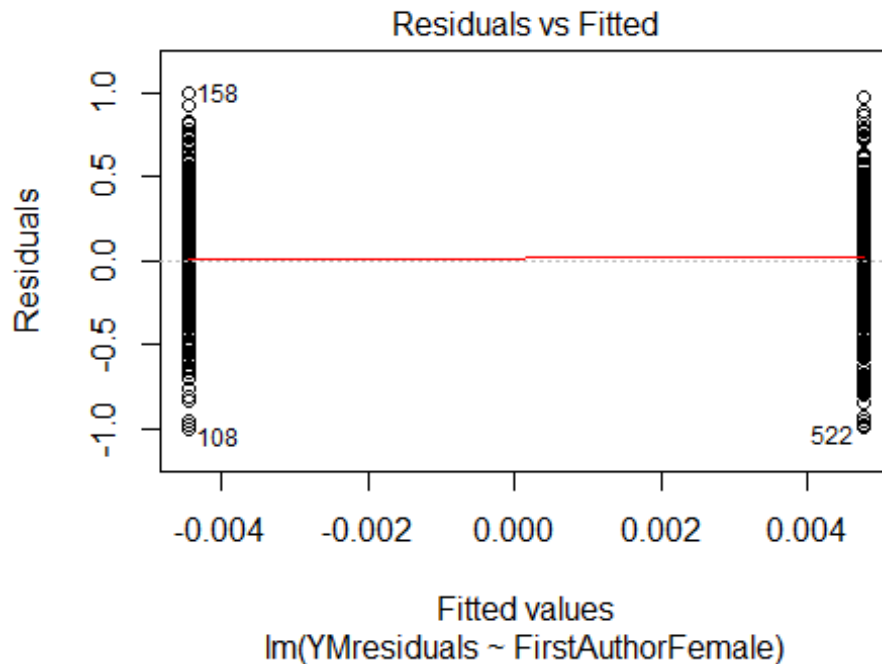
```

##          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: 133"
## [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
##   44   44   42   61   46   37   58   55   63   76   76   80   90   84  102
## 2011 2012
##   92  108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   27   28   37   29   23   45   41   48   57   60   57   70   64   85
## 2011 2012
##   71   88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   28   25   23   29   25   22   42   34   43   49   53   51   64   60   78
## 2011 2012
##   68   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 = 16, df = 16, p-value = 0.4

```

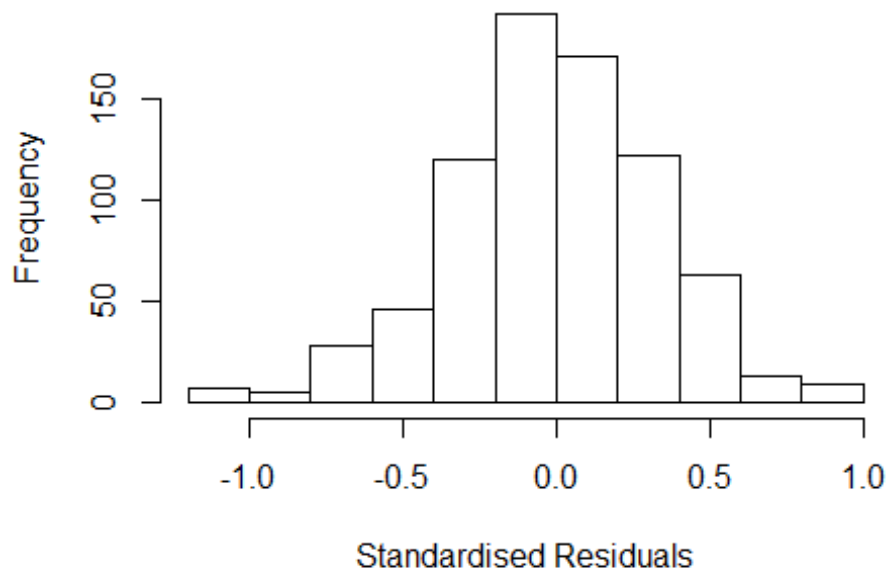


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.59, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 4.31491972432258"
## [1] "Male first author team size 2018 geometric mean: 4.17581902632538"
##
## 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: 3.8942073557186"
## [1] "Male last author team size 2018 geometric mean: 4.56393752261415"
##
## 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 1.211 1          1.100
## LastAuthorFemale  1.161 1          1.077
## UniqueAuthors    1.855 4          1.080
## Year              2.230 16         1.025
```


Residuals from first and last author and team size



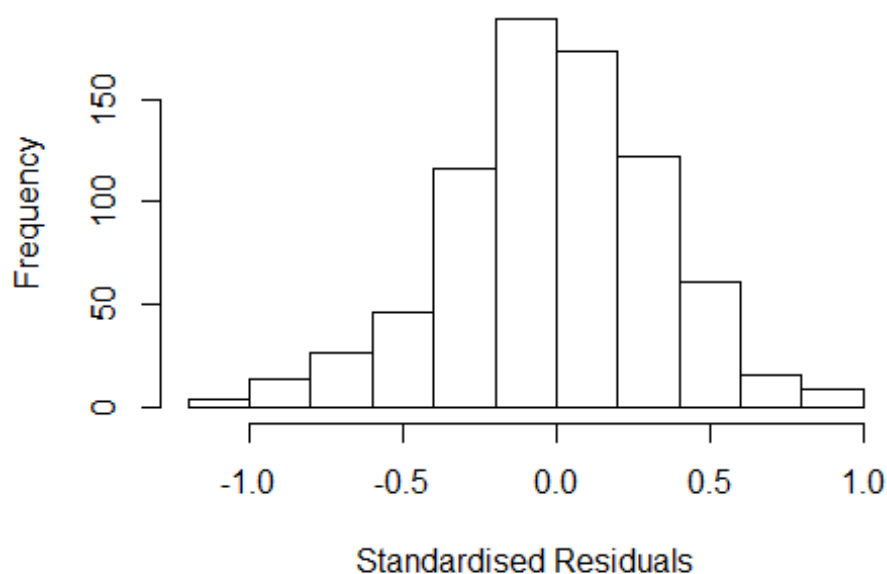
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0575 -0.2198 -0.0139  0.2277  0.9607
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9673    0.1119   8.65 < 2e-16 ***
## FirstAuthorFemale1 -0.0161    0.0259  -0.62  0.5347
## LastAuthorFemale1 -0.0302    0.0279  -1.08  0.2798
## UniqueAuthors2    0.1448    0.0618   2.34  0.0193 *
## UniqueAuthors3    0.1475    0.0617   2.39  0.0170 *
## UniqueAuthors4    0.1671    0.0637   2.62  0.0089 **
## UniqueAuthors5    0.2648    0.0606   4.37 1.4e-05 ***
## Year1997         -0.1188    0.1157  -1.03  0.3051
## Year1998         -0.0450    0.1364  -0.33  0.7413
## Year1999          0.1295    0.1204   1.08  0.2825
```

```

## Year2000          -0.0266      0.1168   -0.23   0.8199
## Year2001          -0.1270      0.1203   -1.06   0.2916
## Year2002          -0.0736      0.1010   -0.73   0.4667
## Year2003          -0.0973      0.1023   -0.95   0.3418
## Year2004          -0.0727      0.1062   -0.68   0.4941
## Year2005          -0.1180      0.1011   -1.17   0.2433
## Year2006          -0.0626      0.1021   -0.61   0.5399
## Year2007          -0.0443      0.1045   -0.42   0.6716
## Year2008          -0.1342      0.0967   -1.39   0.1654
## Year2009          -0.1746      0.0981   -1.78   0.0754 .
## Year2010          -0.1447      0.0985   -1.47   0.1421
## Year2011          -0.0982      0.1011   -0.97   0.3320
## Year2012          -0.1579      0.0993   -1.59   0.1122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.324
## Multiple R-squared:  0.0641, Adjusted R-squared:  0.0368
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 714 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.870  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      1.29e-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 1.174 1      1.083
## LastAuthorFemale  1.129 1      1.063
## Year              1.323 16      1.009

```

Residuals from first and last author



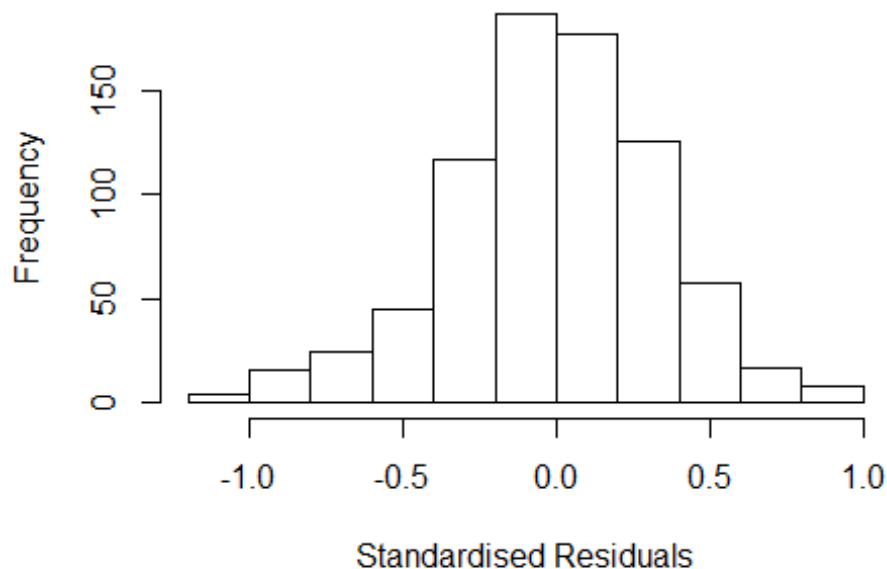
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.04009 -0.21849 -0.00748 0.21890 0.96611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1182 0.0917 12.19 <2e-16 ***
## FirstAuthorFemale1 -0.0023 0.0260 -0.09 0.93
## LastAuthorFemale1 -0.0359 0.0279 -1.29 0.20
## Year1997 -0.1187 0.1172 -1.01 0.31
## Year1998 -0.0718 0.1319 -0.54 0.59
## Year1999 0.0820 0.1173 0.70 0.48
## Year2000 -0.0614 0.1136 -0.54 0.59
## Year2001 -0.1184 0.1225 -0.97 0.33
## Year2002 -0.0900 0.1022 -0.88 0.38
## Year2003 -0.0908 0.1054 -0.86 0.39
## Year2004 -0.0781 0.1068 -0.73 0.46
## Year2005 -0.1115 0.1006 -1.11 0.27
```

```

## Year2006          -0.0731      0.1031   -0.71      0.48
## Year2007          -0.0317      0.1059   -0.30      0.76
## Year2008          -0.1172      0.0984   -1.19      0.23
## Year2009          -0.1426      0.0991   -1.44      0.15
## Year2010          -0.1335      0.0997   -1.34      0.18
## Year2011          -0.0852      0.1027   -0.83      0.41
## Year2012          -0.1503      0.1013   -1.48      0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0255, Adjusted R-squared:  0.00229
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 707 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.294  0.870  0.948  0.897  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      1.29e-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 1.165 1          1.079
## Year              1.165 16          1.005

```

Residuals from first author



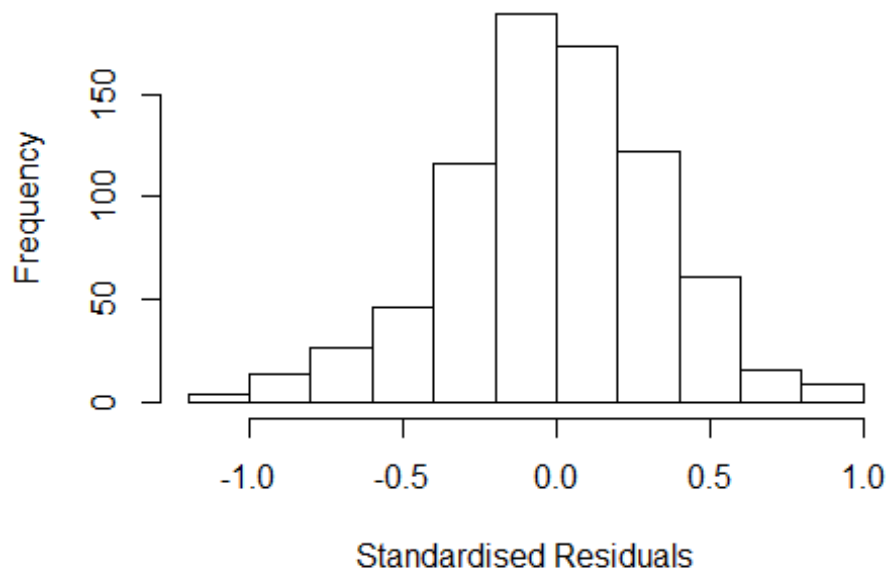
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.02746 -0.22756 -0.00463 0.21737 0.97496
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1006 0.0921 11.95 <2e-16 ***
## FirstAuthorFemale1 -0.0048 0.0261 -0.18 0.85
## Year1997 -0.1159 0.1183 -0.98 0.33
## Year1998 -0.0572 0.1316 -0.43 0.66
## Year1999 0.0921 0.1182 0.78 0.44
## Year2000 -0.0488 0.1145 -0.43 0.67
## Year2001 -0.1077 0.1234 -0.87 0.38
## Year2002 -0.0788 0.1031 -0.76 0.45
## Year2003 -0.0848 0.1064 -0.80 0.43
## Year2004 -0.0731 0.1080 -0.68 0.50
## Year2005 -0.1003 0.1017 -0.99 0.32
## Year2006 -0.0668 0.1046 -0.64 0.52
```

```

## Year2007          -0.0233      0.1067   -0.22      0.83
## Year2008          -0.1138      0.0999   -1.14      0.25
## Year2009          -0.1393      0.1006   -1.38      0.17
## Year2010          -0.1305      0.1013   -1.29      0.20
## Year2011          -0.0789      0.1036   -0.76      0.45
## Year2012          -0.1416      0.1024   -1.38      0.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.000998
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 699 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.312  0.877   0.946   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      1.29e-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 1.124 1          1.060
## Year            1.124 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
## -1.0395 -0.2196 -0.0069 0.2176 0.9675
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1178 0.0918 12.18 <2e-16 ***
## LastAuthorFemale1 -0.0361 0.0280 -1.29 0.20
## Year1997 -0.1187 0.1171 -1.01 0.31
## Year1998 -0.0722 0.1314 -0.55 0.58
## Year1999 0.0817 0.1171 0.70 0.49
## Year2000 -0.0617 0.1133 -0.54 0.59
## Year2001 -0.1185 0.1225 -0.97 0.33
## Year2002 -0.0905 0.1015 -0.89 0.37
## Year2003 -0.0913 0.1052 -0.87 0.39
## Year2004 -0.0783 0.1066 -0.73 0.46
## Year2005 -0.1122 0.0999 -1.12 0.26
## Year2006 -0.0737 0.1022 -0.72 0.47
```

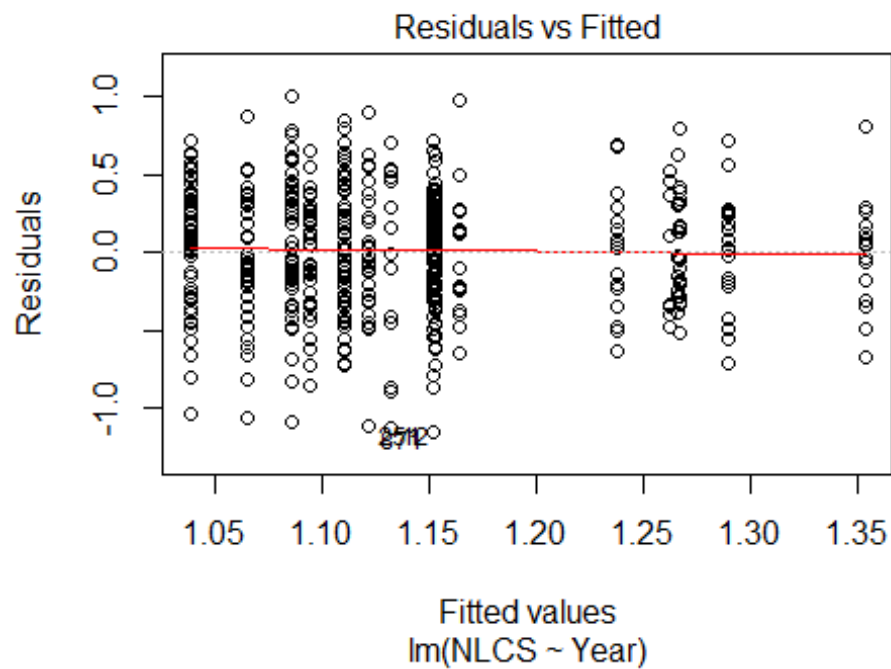
```

## Year2007          -0.0329      0.1046   -0.31      0.75
## Year2008          -0.1181      0.0975   -1.21      0.23
## Year2009          -0.1433      0.0980   -1.46      0.14
## Year2010          -0.1345      0.0984   -1.37      0.17
## Year2011          -0.0861      0.1020   -0.84      0.40
## Year2012          -0.1513      0.0999   -1.52      0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0255, Adjusted R-squared:  0.00363
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 708 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.293  0.870   0.948   0.897   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      1.29e-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: 776"
## [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
##   20   24   16   25   25   15   21   23   32   29   53   69   77   56   67
## 2011 2012
##   81   82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   19    9   19   16   11   16   16   30   27   44   57   66   47   54
## 2011 2012

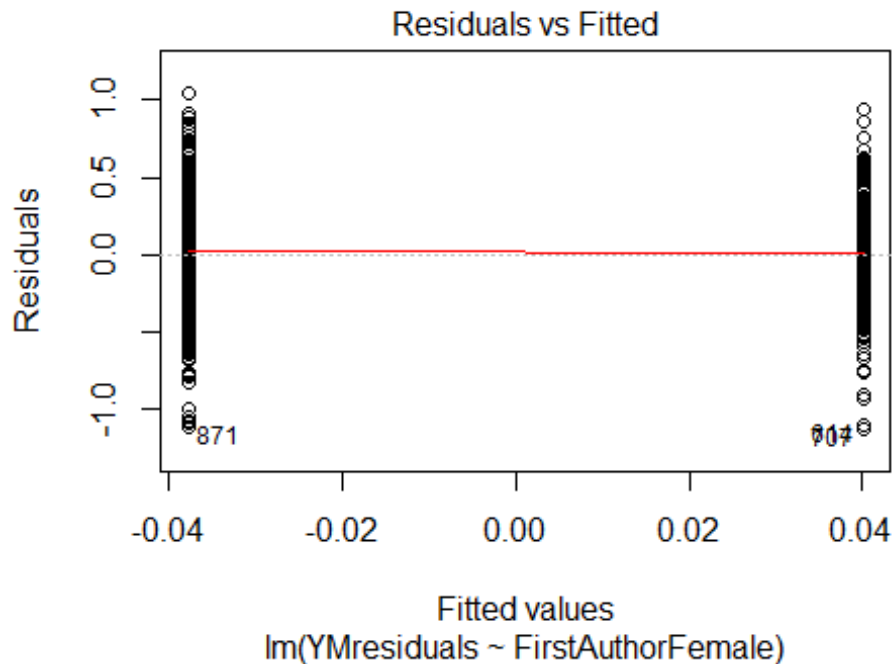
```



```
## 64 54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 19 9 17 13 9 12 14 27 24 37 49 60 44 49
## 2011 2012
## 57 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 = 29, df = 16, p-value = 0.02
```

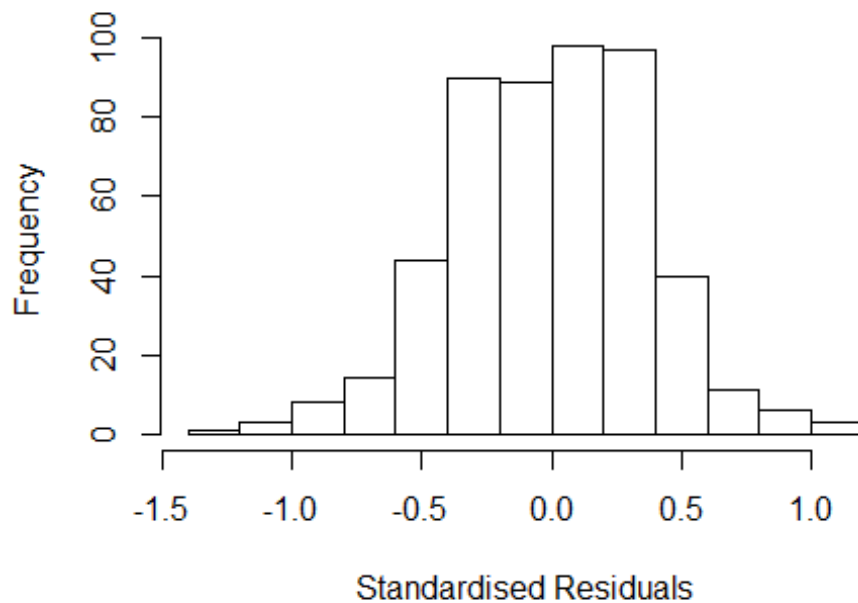


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 6e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 5.21741501049113"
## [1] "Male first author team size 2018 geometric mean: 4.5108507314992"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.44865948091494"
## [1] "Male last author team size 2018 geometric mean: 5.31235914332686"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 900, 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 1.251 1          1.119
## LastAuthorFemale  1.162 1          1.078
## UniqueAuthors    1.911 4          1.084
## Year              2.082 16         1.023
```

Residuals from first and last author and team size



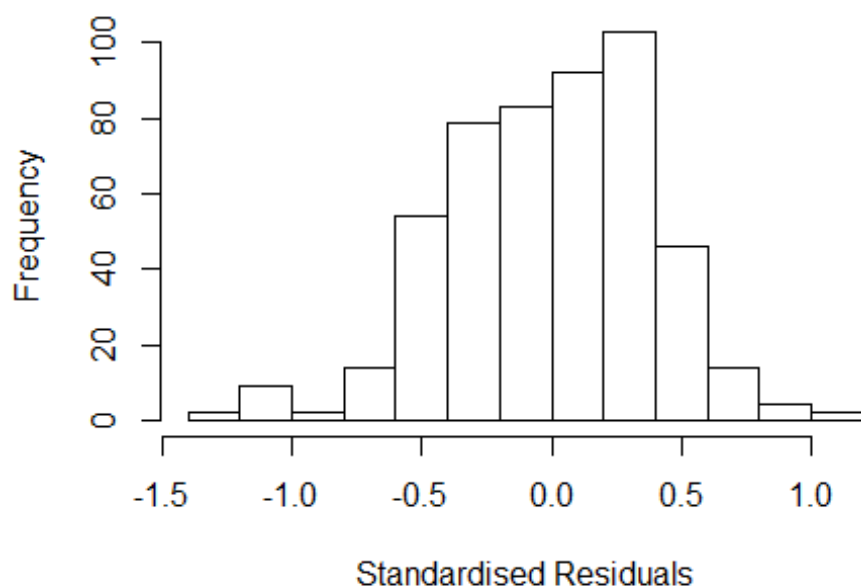
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20722 -0.26146 0.00551 0.25662 1.06239
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9471 0.2370 4.00 7.5e-05 ***
## FirstAuthorFemale1 0.0170 0.0365 0.47 0.64167
## LastAuthorFemale1 0.1183 0.0370 3.20 0.00147 **
## UniqueAuthors2 0.1476 0.1019 1.45 0.14795
## UniqueAuthors3 0.2601 0.0985 2.64 0.00852 **
## UniqueAuthors4 0.3313 0.1023 3.24 0.00128 **
## UniqueAuthors5 0.3750 0.0967 3.88 0.00012 ***
## Year1997 -0.0558 0.2258 -0.25 0.80486
## Year1998 0.0336 0.2862 0.12 0.90664
## Year1999 0.0879 0.2242 0.39 0.69506
```

```

## Year2000          -0.0442      0.2414   -0.18  0.85465
## Year2001           0.0852      0.2275    0.37  0.70826
## Year2002           0.0462      0.2246    0.21  0.83719
## Year2003          -0.0918      0.2382   -0.39  0.70017
## Year2004          -0.1496      0.2268   -0.66  0.50983
## Year2005          -0.0898      0.2173   -0.41  0.67967
## Year2006          -0.1613      0.2209   -0.73  0.46564
## Year2007          -0.1403      0.2193   -0.64  0.52276
## Year2008          -0.1699      0.2200   -0.77  0.44019
## Year2009          -0.1761      0.2211   -0.80  0.42625
## Year2010          -0.1705      0.2233   -0.76  0.44553
## Year2011          -0.0763      0.2187   -0.35  0.72729
## Year2012          -0.1432      0.2187   -0.65  0.51299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.366
## Multiple R-squared:  0.129, Adjusted R-squared:  0.0895
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 467 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.886  0.947  0.906  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      1.98e-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 1.149 1      1.072
## LastAuthorFemale  1.077 1      1.038
## Year              1.224 16      1.006

```

Residuals from first and last author



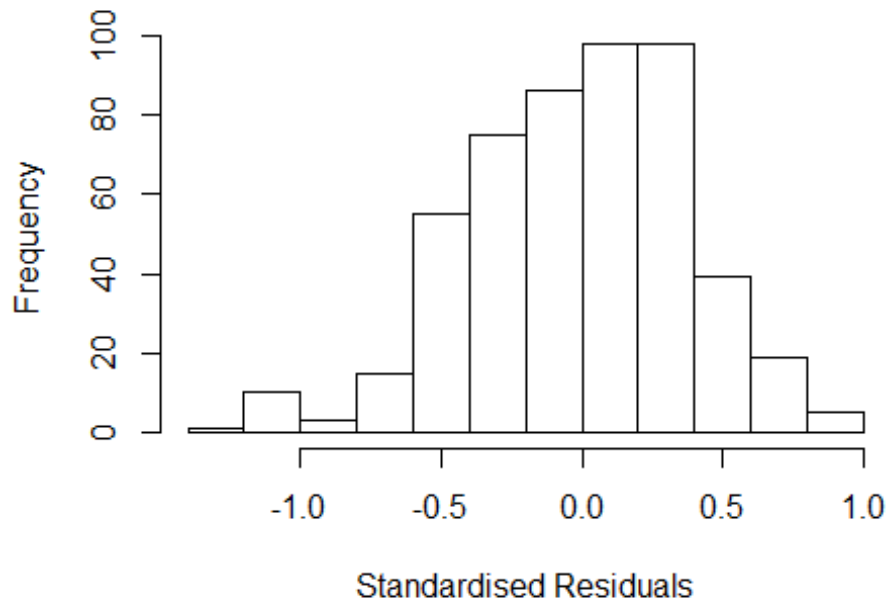
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.238 -0.257 0.016 0.267 1.009
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21788 0.18814 6.47 2.4e-10 ***
## FirstAuthorFemale1 0.05080 0.03673 1.38 0.1673
## LastAuthorFemale1 0.10347 0.03644 2.84 0.0047 **
## Year1997 -0.02837 0.20789 -0.14 0.8915
## Year1998 0.00762 0.24212 0.03 0.9749
## Year1999 0.02671 0.20492 0.13 0.8963
## Year2000 -0.03343 0.23093 -0.14 0.8850
## Year2001 0.06331 0.22032 0.29 0.7739
## Year2002 0.00969 0.20774 0.05 0.9628
## Year2003 -0.13218 0.22704 -0.58 0.5607
## Year2004 -0.18260 0.20591 -0.89 0.3756
## Year2005 -0.09377 0.19861 -0.47 0.6370
```

```

## Year2006          -0.14670      0.19875      -0.74      0.4608
## Year2007          -0.18766      0.20017      -0.94      0.3490
## Year2008          -0.16735      0.19969      -0.84      0.4024
## Year2009          -0.19251      0.20064      -0.96      0.3378
## Year2010          -0.13386      0.20423      -0.66      0.5125
## Year2011          -0.09675      0.19582      -0.49      0.6215
## Year2012          -0.10994      0.19790      -0.56      0.5788
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0474, Adjusted R-squared:  0.012
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 460 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.302  0.886   0.951   0.911   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          1.98e-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 1.146 1      1.070
## Year              1.146 16      1.004

```

Residuals from first author



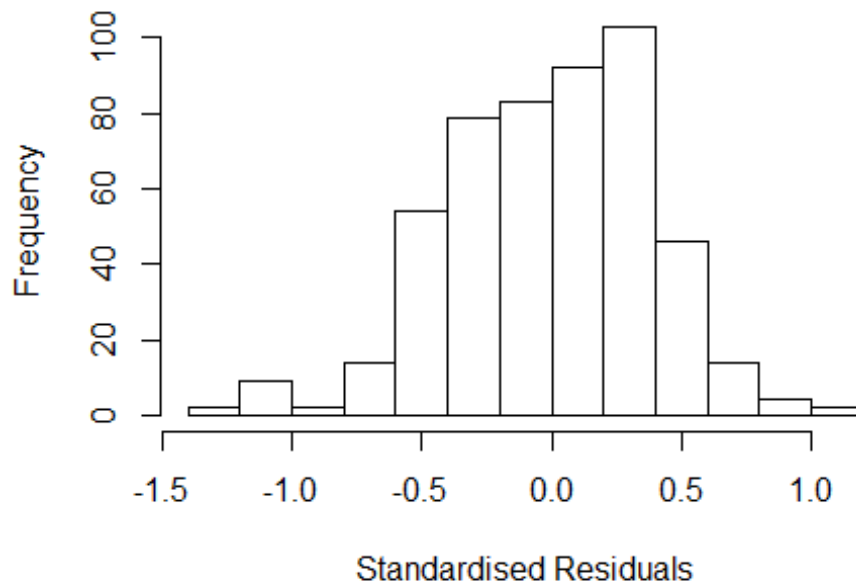
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23889 -0.27316 0.00952 0.27356 0.97363
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.238890 0.199703 6.20 1.2e-09 ***
## FirstAuthorFemale1 0.056871 0.037292 1.53 0.13
## Year1997 0.000541 0.217903 0.00 1.00
## Year1998 -0.007486 0.249863 -0.03 0.98
## Year1999 0.026882 0.215319 0.12 0.90
## Year2000 -0.027310 0.240073 -0.11 0.91
## Year2001 0.052325 0.229244 0.23 0.82
## Year2002 0.003040 0.220199 0.01 0.99
## Year2003 -0.124393 0.236311 -0.53 0.60
## Year2004 -0.189600 0.215540 -0.88 0.38
## Year2005 -0.094827 0.210222 -0.45 0.65
## Year2006 -0.145822 0.209720 -0.70 0.49
```

```

## Year2007          -0.188859    0.212041   -0.89     0.37
## Year2008          -0.158441    0.210537   -0.75     0.45
## Year2009          -0.188620    0.212229   -0.89     0.37
## Year2010          -0.117696    0.214057   -0.55     0.58
## Year2011          -0.094225    0.207193   -0.45     0.65
## Year2012          -0.098231    0.209024   -0.47     0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0335, Adjusted R-squared:  -0.000298
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 460 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.304  0.880  0.951  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      1.98e-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 1.068 1      1.033
## Year              1.068 16      1.002

```


Residuals from last author



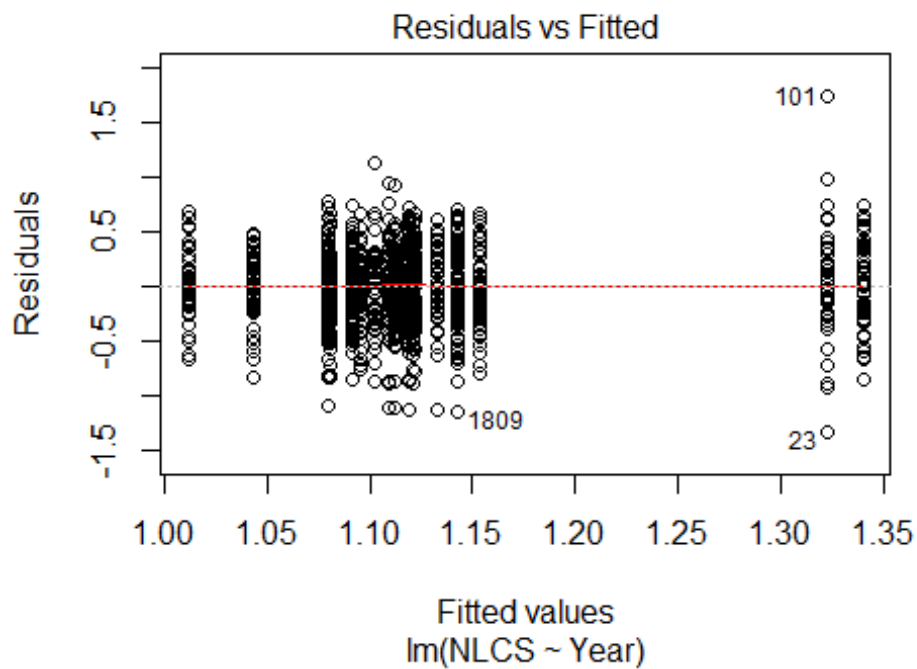
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2206 -0.2649 0.0121 0.2694 1.0459
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22058 0.19048 6.41 3.5e-10 ***
## LastAuthorFemale1 0.10701 0.03669 2.92 0.0037 **
## Year1997 -0.01778 0.20908 -0.09 0.9323
## Year1998 0.03033 0.24616 0.12 0.9020
## Year1999 0.04807 0.20444 0.24 0.8142
## Year2000 -0.00515 0.22578 -0.02 0.9818
## Year2001 0.08637 0.21888 0.39 0.6933
## Year2002 0.03318 0.20819 0.16 0.8734
## Year2003 -0.12145 0.22882 -0.53 0.5958
## Year2004 -0.15971 0.20508 -0.78 0.4365
## Year2005 -0.07019 0.19712 -0.36 0.7220
## Year2006 -0.12736 0.19875 -0.64 0.5219
```

```

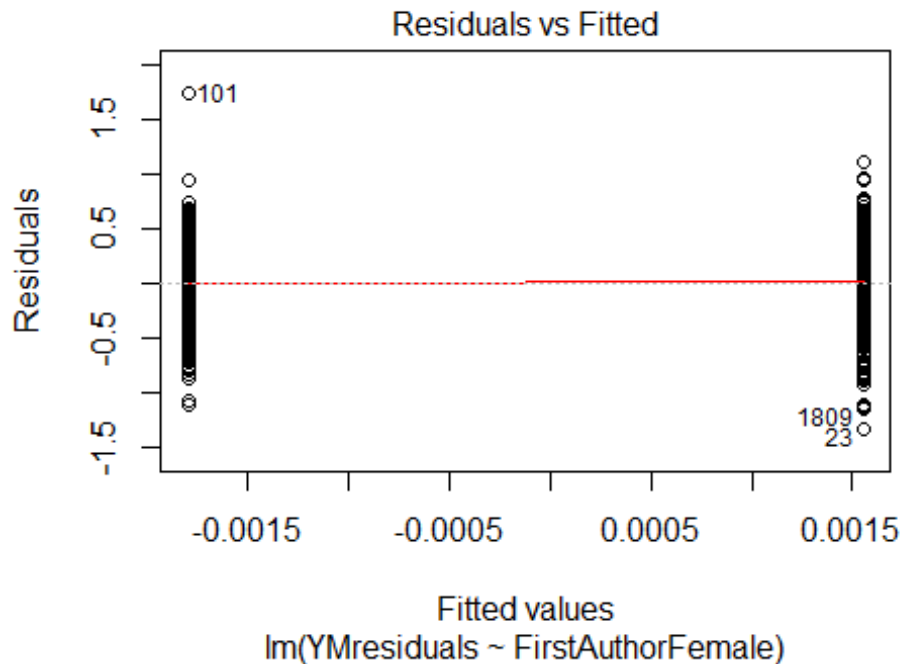
## Year2007          -0.15994      0.19893    -0.80    0.4218
## Year2008          -0.14918      0.19965    -0.75    0.4553
## Year2009          -0.17055      0.19947    -0.85    0.3930
## Year2010          -0.11147      0.20284    -0.55    0.5829
## Year2011          -0.06951      0.19460    -0.36    0.7211
## Year2012          -0.08658      0.19725    -0.44    0.6609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0425, Adjusted R-squared:  0.00901
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 471 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.315  0.888  0.954  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      1.98e-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: 504"
## [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
## 111  93  89  66  94  102  78  74  96  73  91  96  111  138  173
## 2011 2012
## 160 171
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41  51  60  44  50  51  57  49  60  47  63  64  77  84  122
## 2011 2012

```

```
## 112 103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 39 46 54 36 44 46 51 42 49 38 52 55 64 69 102
## 2011 2012
## 90 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 = 43, df = 16, p-value = 3e-04
```

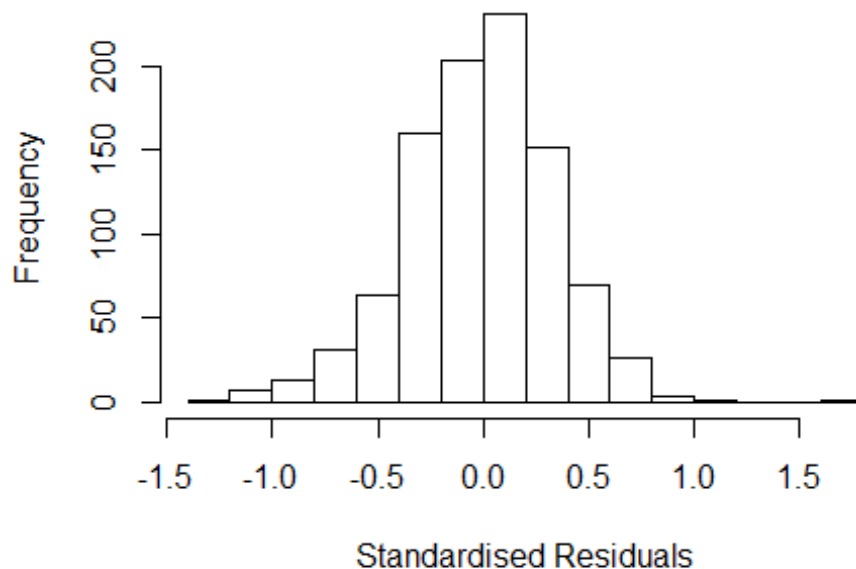


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.3, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 5.50751519928129"
## [1] "Male first author team size 2018 geometric mean: 4.79793730123235"
##
## 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: 5.33907209716886"
## [1] "Male last author team size 2018 geometric mean: 5.08891971283876"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, 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 1.095 1      1.046
## LastAuthorFemale  1.081 1      1.040
## UniqueAuthors    1.416 4      1.044
## Year             1.521 16      1.013
```

Residuals from first and last author and team size



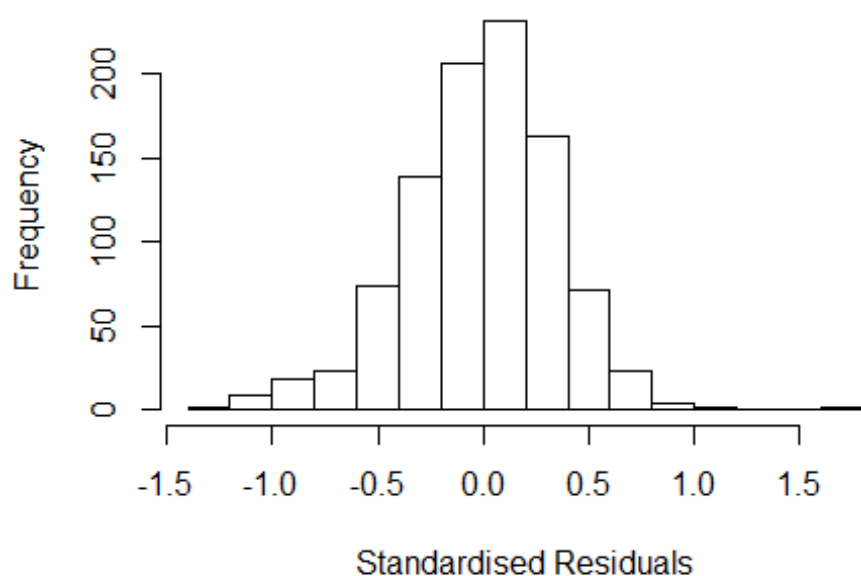
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27794 -0.22386  0.00127  0.20946  1.75968
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28559    0.10844   11.85  < 2e-16 ***
## FirstAuthorFemale1 -0.00904    0.02223   -0.41  0.68426
## LastAuthorFemale1 -0.03842    0.02590   -1.48  0.13839
## UniqueAuthors2    -0.06745    0.09188   -0.73  0.46305
## UniqueAuthors3     0.03076    0.09038    0.34  0.73366
## UniqueAuthors4     0.03403    0.09106    0.37  0.70870
## UniqueAuthors5     0.08815    0.08946    0.99  0.32472
## Year1997          0.02055    0.09220    0.22  0.82367
## Year1998         -0.16065    0.08979   -1.79  0.07391 .
## Year1999         -0.23153    0.08747   -2.65  0.00826 **
```

```

## Year2000      -0.18424      0.08915      -2.07      0.03903 *
## Year2001      -0.28899      0.08451      -3.42      0.00065 ***
## Year2002      -0.14937      0.08819      -1.69      0.09066 .
## Year2003      -0.20049      0.08595      -2.33      0.01988 *
## Year2004      -0.20287      0.08271      -2.45      0.01435 *
## Year2005      -0.21979      0.09410      -2.34      0.01971 *
## Year2006      -0.17858      0.08511      -2.10      0.03615 *
## Year2007      -0.22135      0.08391      -2.64      0.00848 **
## Year2008      -0.22339      0.08305      -2.69      0.00727 **
## Year2009      -0.18564      0.08206      -2.26      0.02391 *
## Year2010      -0.17685      0.08032      -2.20      0.02792 *
## Year2011      -0.22857      0.08081      -2.83      0.00478 **
## Year2012      -0.15784      0.08586      -1.84      0.06633 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.326
## Multiple R-squared:  0.0605, Adjusted R-squared:  0.0385
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 28 is an outlier with |weight| = 0 ( < 0.0001);
## 81 weights are ~= 1. The remaining 881 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0896 0.8730 0.9520 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           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 1.053 1 1.026
## LastAuthorFemale 1.063 1 1.031
## Year 1.104 16 1.003

```

Residuals from first and last author



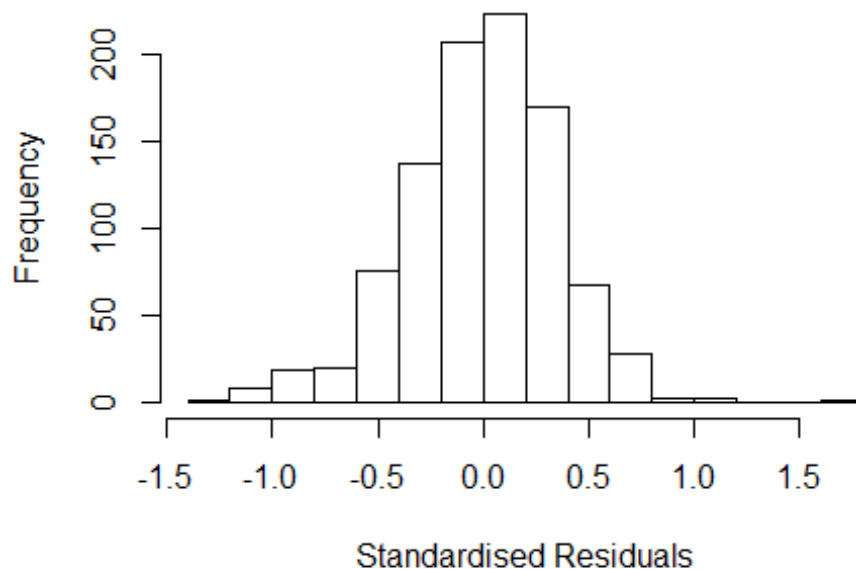
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2639 -0.2176 0.0103 0.2223 1.7649
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30484 0.07689 16.97 < 2e-16 ***
## FirstAuthorFemale1 -0.00273 0.02241 -0.12 0.90310
## LastAuthorFemale1 -0.04090 0.02640 -1.55 0.12164
## Year1997 0.01346 0.09471 0.14 0.88699
## Year1998 -0.16731 0.09113 -1.84 0.06670 .
## Year1999 -0.21632 0.08974 -2.41 0.01612 *
## Year2000 -0.17123 0.09030 -1.90 0.05822 .
## Year2001 -0.29038 0.08784 -3.31 0.00098 ***
## Year2002 -0.15114 0.09027 -1.67 0.09440 .
## Year2003 -0.19073 0.08860 -2.15 0.03159 *
## Year2004 -0.19135 0.08540 -2.24 0.02528 *
## Year2005 -0.19383 0.09818 -1.97 0.04865 *
```

```

## Year2006      -0.16009    0.08779   -1.82  0.06851 .
## Year2007      -0.20734    0.08559   -2.42  0.01560 *
## Year2008      -0.20299    0.08504   -2.39  0.01719 *
## Year2009      -0.18416    0.08488   -2.17  0.03029 *
## Year2010      -0.15981    0.08183   -1.95  0.05111 .
## Year2011      -0.20606    0.08247   -2.50  0.01264 *
## Year2012      -0.12969    0.08819   -1.47  0.14177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0371, Adjusted R-squared:  0.0187
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 28 is an outlier with |weight| = 0 ( < 0.0001);
## 83 weights are ~= 1. The remaining 879 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.105  0.867   0.950   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           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 1.043  1         1.021
## Year              1.043 16         1.001

```


Residuals from first author



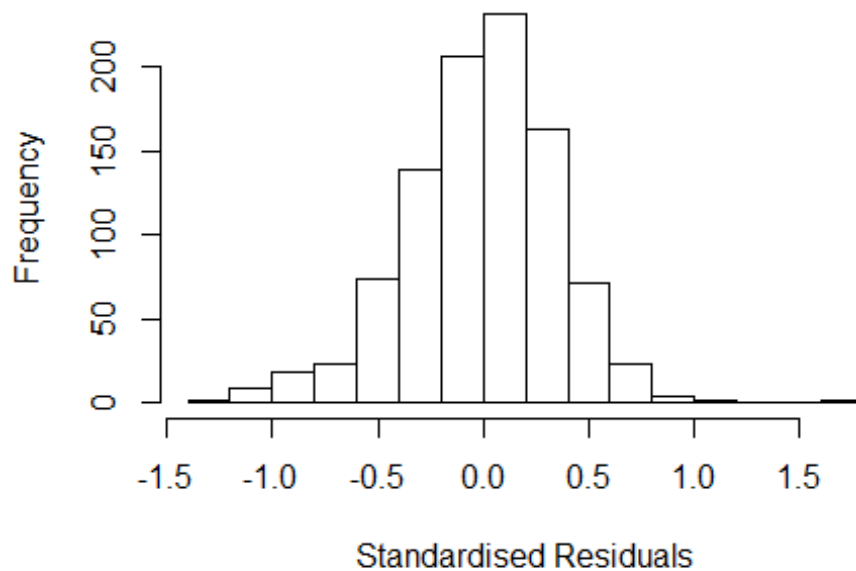
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29647 -0.21921 0.00794 0.22101 1.77702
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29647 0.07544 17.19 < 2e-16 ***
## FirstAuthorFemale1 -0.00649 0.02229 -0.29 0.77105
## Year1997 0.01550 0.09329 0.17 0.86809
## Year1998 -0.16220 0.09001 -1.80 0.07187 .
## Year1999 -0.21232 0.08886 -2.39 0.01707 *
## Year2000 -0.17492 0.08974 -1.95 0.05156 .
## Year2001 -0.29154 0.08662 -3.37 0.00079 ***
## Year2002 -0.14752 0.08932 -1.65 0.09895 .
## Year2003 -0.18749 0.08772 -2.14 0.03282 *
## Year2004 -0.19086 0.08406 -2.27 0.02340 *
## Year2005 -0.19189 0.09716 -1.97 0.04856 *
## Year2006 -0.15399 0.08659 -1.78 0.07567 .
```

```

## Year2007          -0.20799    0.08460    -2.46    0.01413 *
## Year2008          -0.20022    0.08388    -2.39    0.01718 *
## Year2009          -0.18369    0.08360    -2.20    0.02825 *
## Year2010          -0.16195    0.08070    -2.01    0.04506 *
## Year2011          -0.20756    0.08133    -2.55    0.01087 *
## Year2012          -0.13031    0.08700    -1.50    0.13450
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0349, Adjusted R-squared:  0.0175
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 28 is an outlier with |weight| = 0 ( < 0.0001);
## 81 weights are ~= 1. The remaining 881 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0827 0.8680 0.9500 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      1.04e-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 1.055 1          1.027
## Year            1.055 16          1.002

```

Residuals from last author



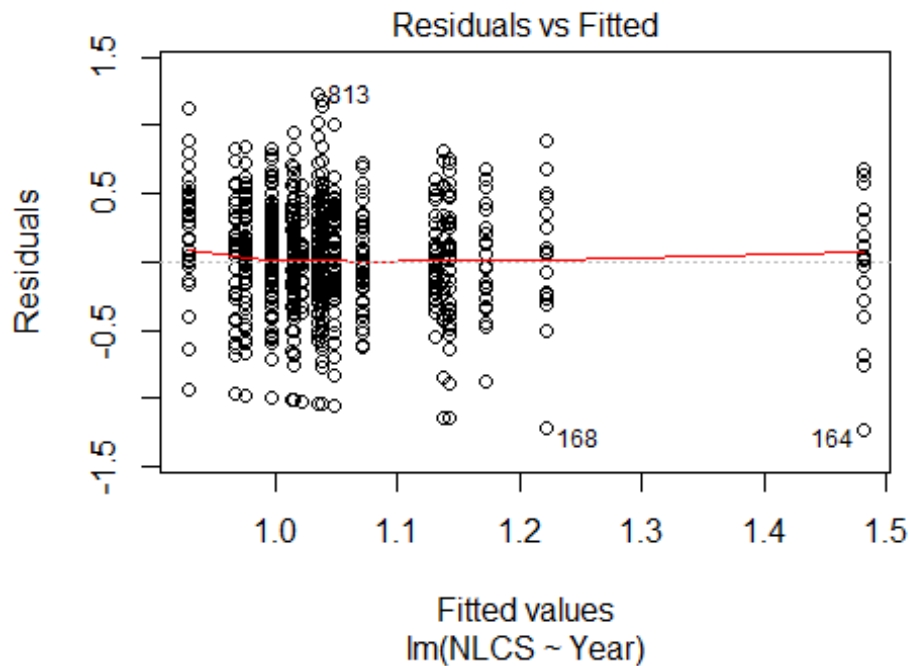
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2621 -0.2176 0.0114 0.2235 1.7636
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3034 0.0753 17.31 < 2e-16 ***
## LastAuthorFemale1 -0.0412 0.0263 -1.57 0.11723
## Year1997 0.0136 0.0947 0.14 0.88543
## Year1998 -0.1671 0.0910 -1.84 0.06681 .
## Year1999 -0.2160 0.0897 -2.41 0.01621 *
## Year2000 -0.1707 0.0900 -1.90 0.05822 .
## Year2001 -0.2899 0.0877 -3.31 0.00098 ***
## Year2002 -0.1511 0.0903 -1.67 0.09449 .
## Year2003 -0.1907 0.0886 -2.15 0.03163 *
## Year2004 -0.1911 0.0852 -2.24 0.02516 *
## Year2005 -0.1935 0.0982 -1.97 0.04916 *
## Year2006 -0.1600 0.0878 -1.82 0.06877 .
```

```

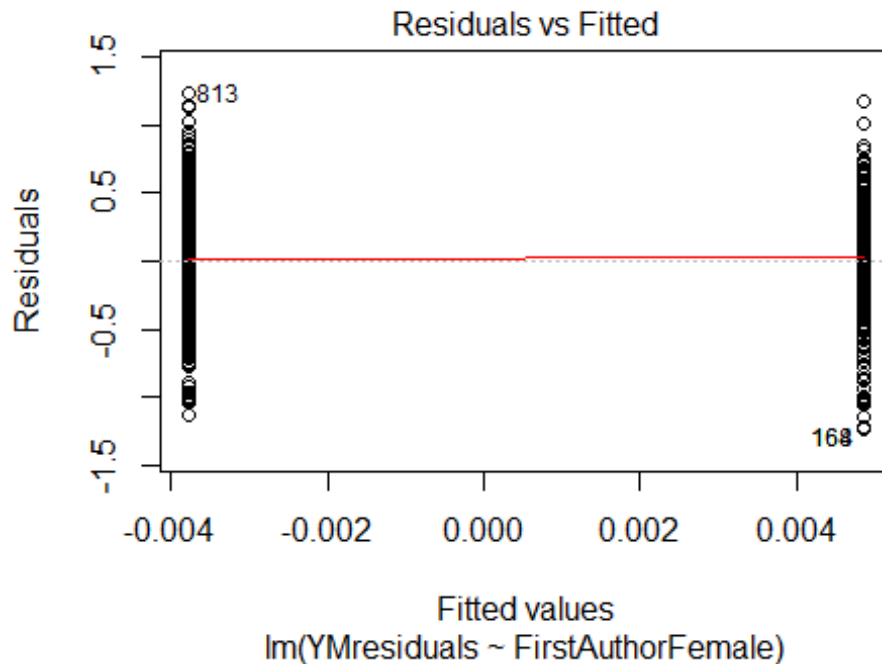
## Year2007          -0.2072      0.0855   -2.42  0.01559 *
## Year2008          -0.2027      0.0849   -2.39  0.01716 *
## Year2009          -0.1840      0.0848   -2.17  0.03034 *
## Year2010          -0.1598      0.0819   -1.95  0.05120 .
## Year2011          -0.2059      0.0824   -2.50  0.01266 *
## Year2012          -0.1293      0.0880   -1.47  0.14231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.037, Adjusted R-squared:  0.0197
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 28 is an outlier with |weight| = 0 ( < 0.0001);
## 81 weights are ~= 1. The remaining 881 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.867  0.951  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      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: 963"
## [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
##   40   26   34   31   30   32   45   38   55   53   74   60   74   96  120
## 2011 2012
##   99  125
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32   20   23   21   18   14   35   27   39   41   56   41   63   75   99

```

```
## 2011 2012
## 75 99
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 19 22 19 17 12 32 24 32 33 49 40 55 71 91
## 2011 2012
## 69 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 = 37, df = 16, p-value = 0.002
```

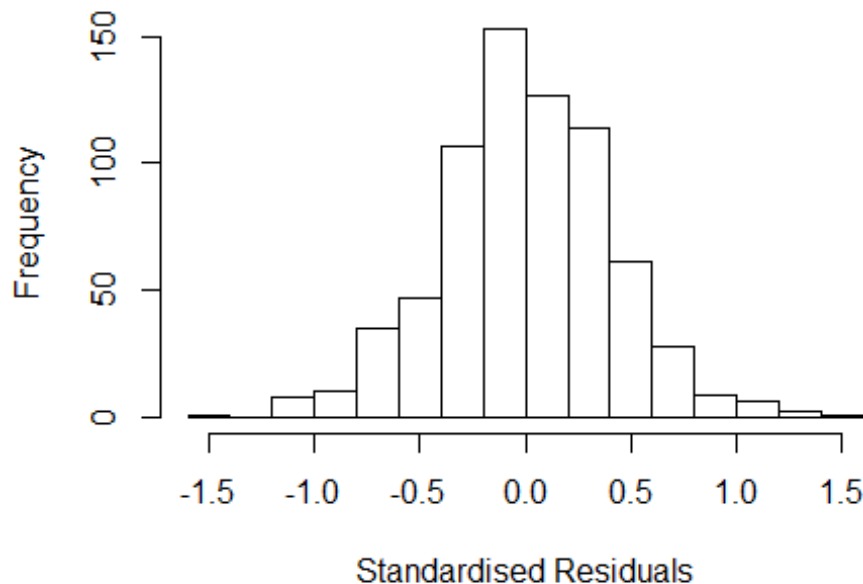


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.94, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.89334765751339"
## [1] "Male first author team size 2018 geometric mean: 3.82340847421115"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.59954566676471"
## [1] "Male last author team size 2018 geometric mean: 4.09027180126815"
##
## 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 1.159 1      1.077
## LastAuthorFemale  1.159 1      1.077
## UniqueAuthors    1.807 4      1.077
## Year              2.044 16     1.023
```

Residuals from first and last author and team size



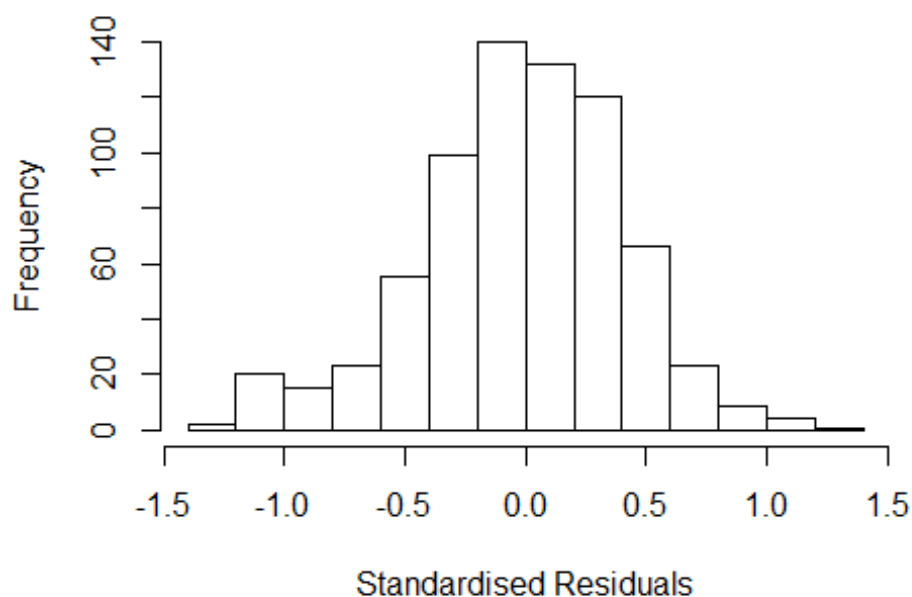
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4607 -0.2471 -0.0105 0.2603 1.4724
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9832 0.1049 9.37 < 2e-16 ***
## FirstAuthorFemale1 -0.0385 0.0317 -1.22 0.22424
## LastAuthorFemale1 -0.0436 0.0340 -1.28 0.20083
## UniqueAuthors2 0.2793 0.0744 3.75 0.00019 ***
## UniqueAuthors3 0.3062 0.0709 4.32 1.8e-05 ***
## UniqueAuthors4 0.3196 0.0755 4.24 2.6e-05 ***
## UniqueAuthors5 0.3845 0.0707 5.44 7.4e-08 ***
## Year1997 -0.0374 0.1403 -0.27 0.78996
## Year1998 0.0080 0.1300 0.06 0.95093
## Year1999 0.3579 0.1523 2.35 0.01905 *
```

```

## Year2000          0.0930      0.1630      0.57  0.56846
## Year2001          -0.1286      0.2029     -0.63  0.52635
## Year2002          -0.0391      0.1094     -0.36  0.72068
## Year2003          -0.2105      0.1296     -1.62  0.10482
## Year2004          -0.2661      0.1232     -2.16  0.03123 *
## Year2005          -0.1668      0.1119     -1.49  0.13633
## Year2006          -0.1476      0.1085     -1.36  0.17400
## Year2007          -0.2039      0.1366     -1.49  0.13588
## Year2008          -0.2808      0.1121     -2.51  0.01246 *
## Year2009          -0.2438      0.1052     -2.32  0.02076 *
## Year2010          -0.1925      0.1083     -1.78  0.07586 .
## Year2011          -0.2461      0.1060     -2.32  0.02051 *
## Year2012          -0.2311      0.1073     -2.15  0.03165 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.144, Adjusted R-squared:  0.116
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 640 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.116  0.866  0.954  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          1.41e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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  1.209 1          1.100
## Year              1.286 16          1.008

```


Residuals from first and last author



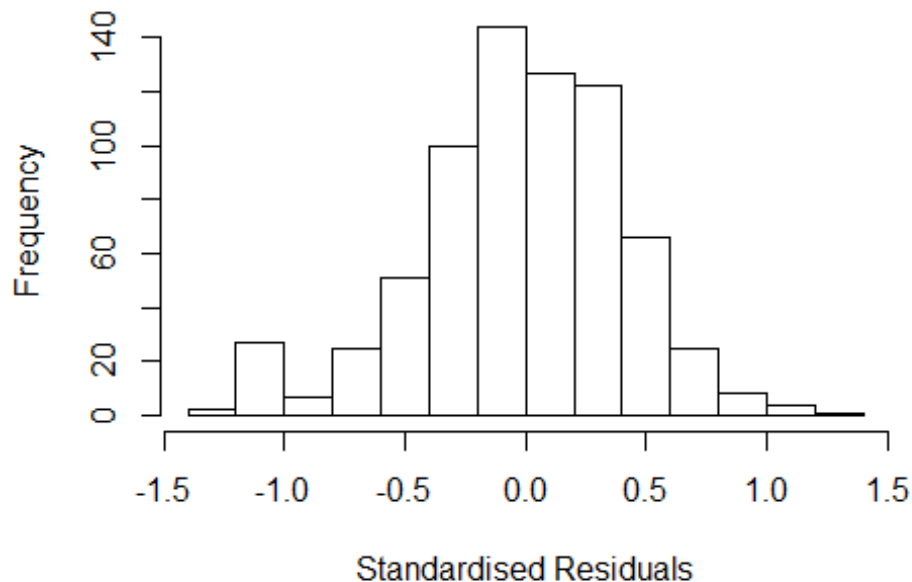
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.306164 -0.251304  0.000309  0.263584  1.284423
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2063     0.1055   11.43  <2e-16 ***
## FirstAuthorFemale1 -0.0064     0.0319   -0.20    0.841
## LastAuthorFemale1 -0.0500     0.0359   -1.39    0.164
## Year1997          0.0144     0.1490    0.10    0.923
## Year1998          0.0106     0.1369    0.08    0.939
## Year1999          0.3559     0.1533    2.32    0.021 *
## Year2000          0.0822     0.1587    0.52    0.605
## Year2001         -0.0531     0.1975   -0.27    0.788
## Year2002         -0.0482     0.1181   -0.41    0.683
## Year2003         -0.1376     0.1353   -1.02    0.310
## Year2004         -0.2317     0.1351   -1.72    0.087 .
## Year2005         -0.0952     0.1213   -0.78    0.433
```

```

## Year2006          -0.1091      0.1184    -0.92      0.357
## Year2007          -0.1609      0.1653    -0.97      0.331
## Year2008          -0.2210      0.1210    -1.83      0.068 .
## Year2009          -0.1753      0.1141    -1.54      0.125
## Year2010          -0.1301      0.1165    -1.12      0.264
## Year2011          -0.1775      0.1156    -1.54      0.125
## Year2012          -0.1620      0.1150    -1.41      0.159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0683, Adjusted R-squared:  0.044
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 644 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.229  0.857  0.949  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      1.41e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



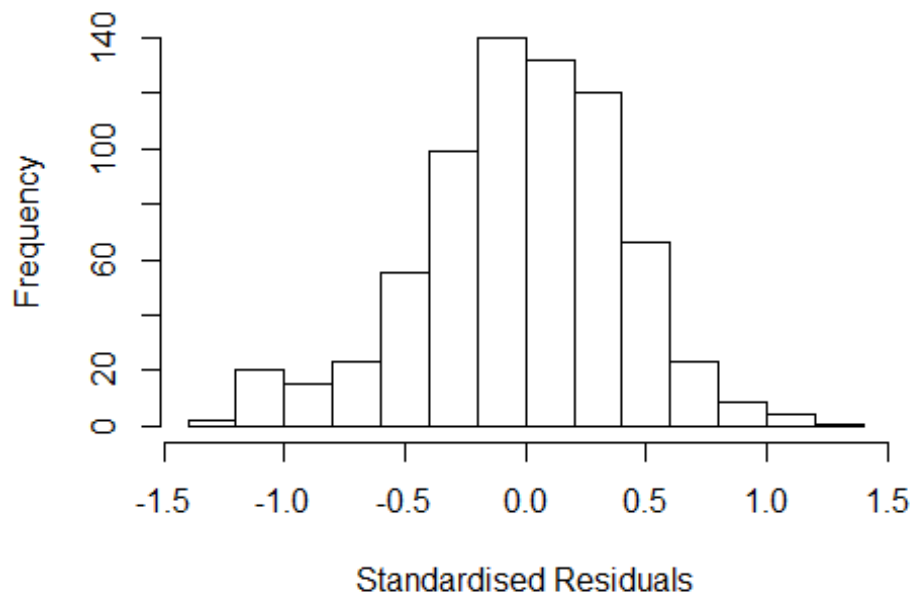
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28858 -0.25582 -0.00734 0.27237 1.25540
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19463 0.10731 11.13 <2e-16 ***
## FirstAuthorFemale1 -0.01536 0.03202 -0.48 0.632
## Year1997 0.00999 0.15065 0.07 0.947
## Year1998 0.01184 0.13715 0.09 0.931
## Year1999 0.34994 0.15425 2.27 0.024 *
## Year2000 0.08140 0.15846 0.51 0.608
## Year2001 -0.05641 0.19517 -0.29 0.773
## Year2002 -0.04222 0.11915 -0.35 0.723
## Year2003 -0.12818 0.13510 -0.95 0.343
## Year2004 -0.23289 0.13726 -1.70 0.090 .
## Year2005 -0.09399 0.12295 -0.76 0.445
## Year2006 -0.11591 0.12005 -0.97 0.335
```

```

## Year2007          -0.15390    0.16661   -0.92    0.356
## Year2008          -0.22601    0.12274   -1.84    0.066 .
## Year2009          -0.17568    0.11593   -1.52    0.130
## Year2010          -0.13345    0.11844   -1.13    0.260
## Year2011          -0.17593    0.11739   -1.50    0.134
## Year2012          -0.16349    0.11684   -1.40    0.162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.0641, Adjusted R-squared:  0.041
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 645 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.245  0.862  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      1.41e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.197 1          1.094
## Year            1.197 16          1.006

```

Residuals from last author



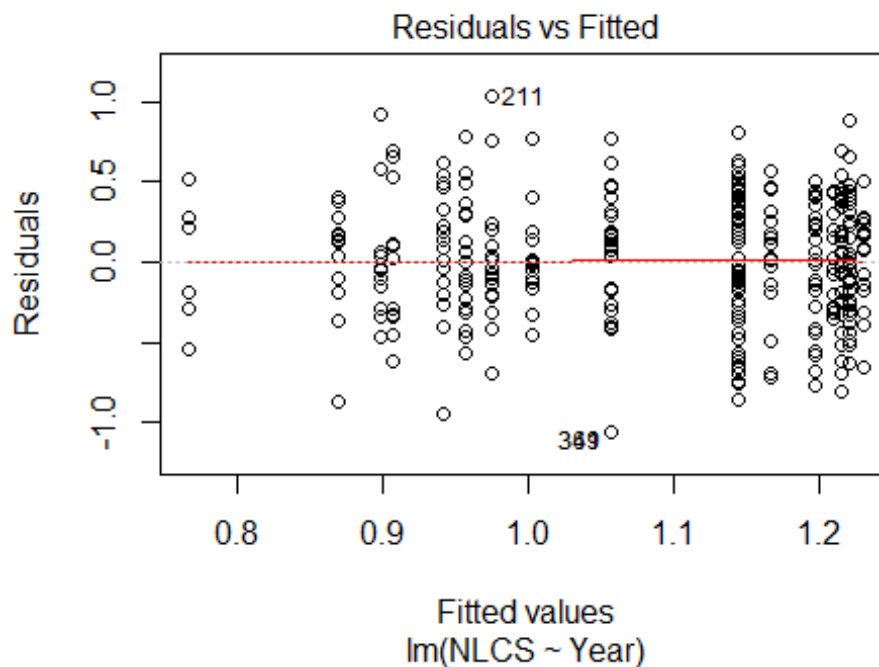
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30289 -0.24951  0.00316  0.26269  1.28316
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2043     0.1057   11.39  <2e-16 ***
## LastAuthorFemale1 -0.0513     0.0359   -1.43    0.153
## Year1997          0.0146     0.1490    0.10    0.922
## Year1998          0.0105     0.1367    0.08    0.939
## Year1999          0.3546     0.1524    2.33    0.020 *
## Year2000          0.0818     0.1583    0.52    0.606
## Year2001         -0.0533     0.1973   -0.27    0.787
## Year2002         -0.0490     0.1175   -0.42    0.677
## Year2003         -0.1391     0.1343   -1.04    0.300
## Year2004         -0.2321     0.1348   -1.72    0.086 .
## Year2005         -0.0966     0.1205   -0.80    0.423
## Year2006         -0.1098     0.1177   -0.93    0.351
```

```

## Year2007          -0.1615      0.1645   -0.98      0.326
## Year2008          -0.2226      0.1199   -1.86      0.064 .
## Year2009          -0.1771      0.1128   -1.57      0.117
## Year2010          -0.1313      0.1154   -1.14      0.256
## Year2011          -0.1792      0.1144   -1.57      0.118
## Year2012          -0.1628      0.1144   -1.42      0.155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0684, Adjusted R-squared:  0.0454
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.858  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      1.41e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 709"
## [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
##   13   25   16   26   22   31   19   27   23   22   22   33   36   38   37
## 2011 2012
##   42   42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6   13   13   19   19   16   14   18   13   17   20   25   28   32   29
## 2011 2012

```

```
## 31 39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 13 11 17 17 14 12 14 10 16 19 24 27 27 28
## 2011 2012
## 30 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 = 9.7, df = 16, p-value = 0.9
```



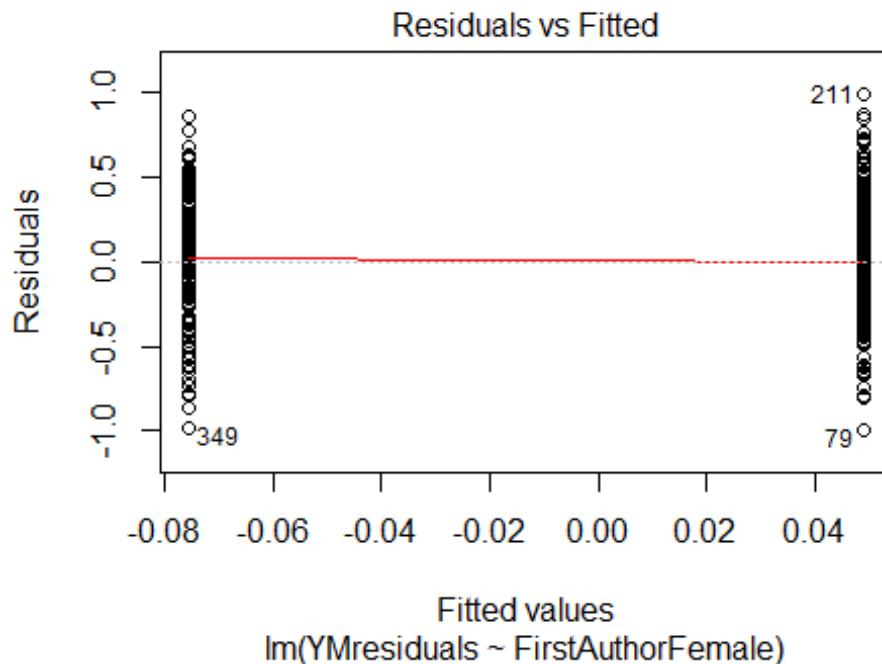
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.5, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 4.66467527591722"
## [1] "Male first author team size 2018 geometric mean: 4.00779865332646"

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

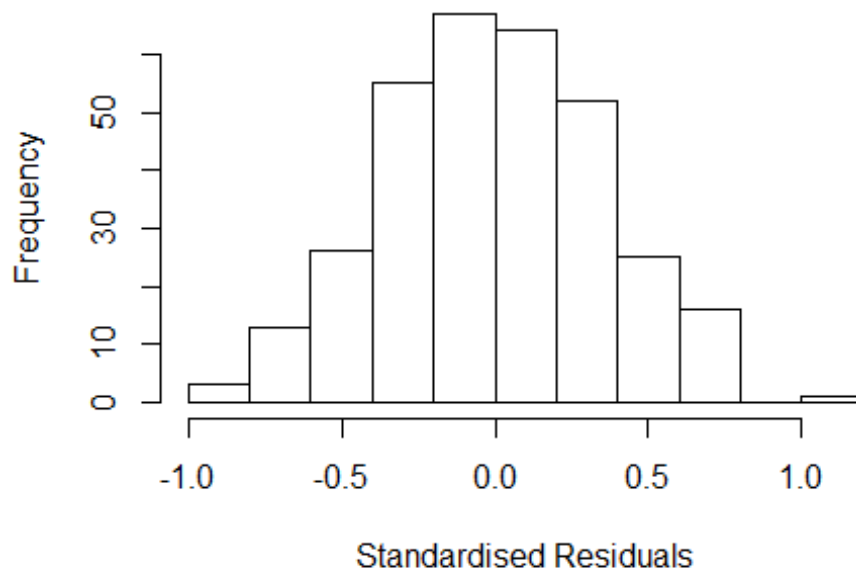
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.37711771467215"
## [1] "Male last author team size 2018 geometric mean: 3.71324919740096"

## 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 = 0.04
## 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 1.283 1      1.133
## LastAuthorFemale  1.723 1      1.313
## UniqueAuthors    2.935 4      1.144
## Year              4.937 16     1.051
```


Residuals from first and last author and team size



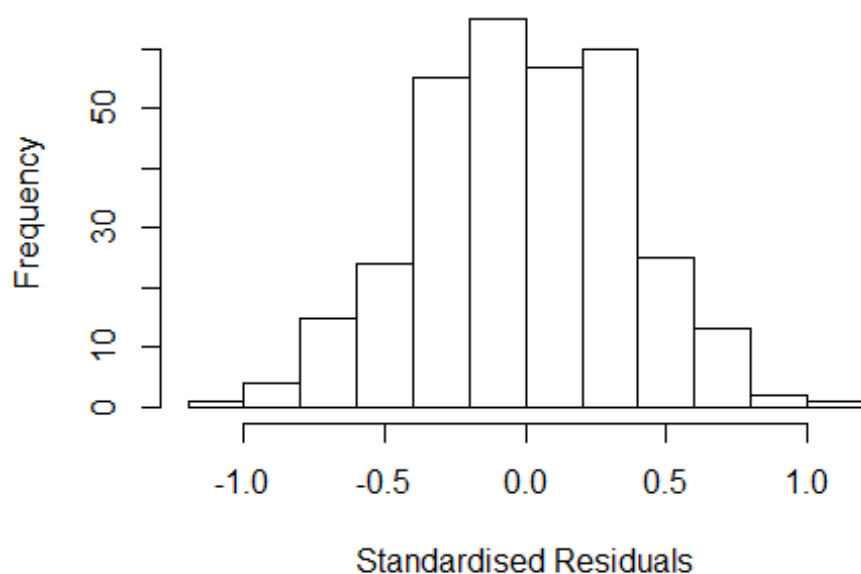
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.98341 -0.23860 -0.00677 0.25322 1.07202
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4344 0.1811 2.40 0.0170 *
## FirstAuthorFemale1 0.0607 0.0461 1.32 0.1883
## LastAuthorFemale1 0.0753 0.0494 1.52 0.1285
## UniqueAuthors2 0.2608 0.1215 2.15 0.0326 *
## UniqueAuthors3 0.3201 0.1189 2.69 0.0075 **
## UniqueAuthors4 0.2781 0.1287 2.16 0.0315 *
## UniqueAuthors5 0.3705 0.1163 3.19 0.0016 **
## Year1997 0.0789 0.2238 0.35 0.7245
## Year1998 0.1312 0.1994 0.66 0.5113
## Year1999 0.1522 0.1976 0.77 0.4418
```

```

## Year2000          0.1305      0.1990      0.66      0.5125
## Year2001          0.1762      0.1921      0.92      0.3598
## Year2002          0.1265      0.1895      0.67      0.5050
## Year2003          0.1217      0.1967      0.62      0.5366
## Year2004          0.3331      0.1984      1.68      0.0941 .
## Year2005          0.3971      0.1843      2.15      0.0320 *
## Year2006          0.3609      0.1840      1.96      0.0508 .
## Year2007          0.2600      0.1777      1.46      0.1443
## Year2008          0.3825      0.1847      2.07      0.0392 *
## Year2009          0.3322      0.1788      1.86      0.0641 .
## Year2010          0.4180      0.1803      2.32      0.0211 *
## Year2011          0.3895      0.1812      2.15      0.0323 *
## Year2012          0.3530      0.1745      2.02      0.0440 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.365
## Multiple R-squared:  0.17,   Adjusted R-squared:  0.109
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 291 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.368  0.871  0.950  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          3.11e-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 1.245 1      1.116
## LastAuthorFemale  1.503 1      1.226
## Year              1.800 16      1.019

```

Residuals from first and last author



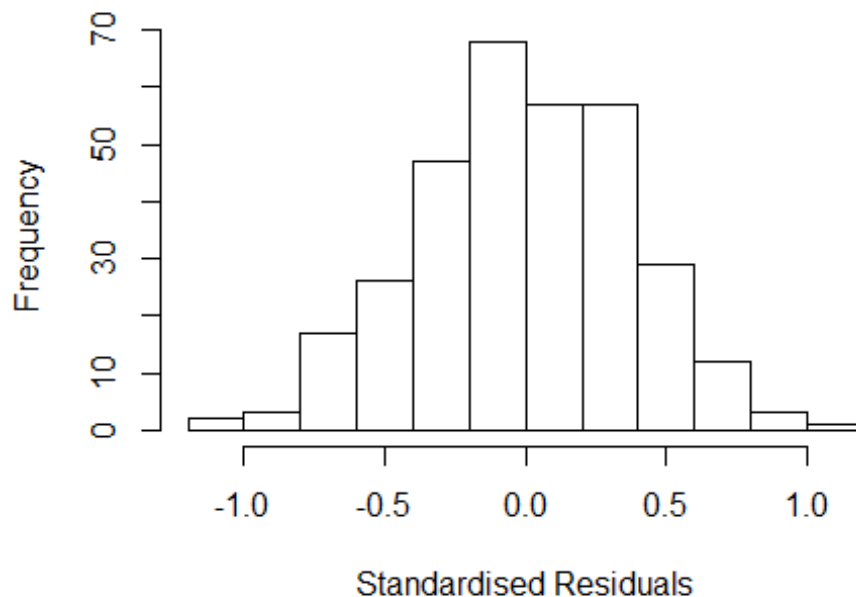
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.03457 -0.25988 -0.00484 0.26362 1.07004
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6451 0.1767 3.65 0.00031 ***
## FirstAuthorFemale1 0.0935 0.0475 1.97 0.04997 *
## LastAuthorFemale1 0.0838 0.0482 1.74 0.08332 .
## Year1997 0.1108 0.2393 0.46 0.64366
## Year1998 0.1832 0.1937 0.95 0.34493
## Year1999 0.2122 0.1987 1.07 0.28633
## Year2000 0.2030 0.1907 1.06 0.28808
## Year2001 0.2308 0.1865 1.24 0.21692
## Year2002 0.1933 0.1827 1.06 0.29085
## Year2003 0.2004 0.1940 1.03 0.30241
## Year2004 0.4280 0.1905 2.25 0.02540 *
## Year2005 0.4861 0.1837 2.65 0.00857 **
```

```

## Year2006          0.4183      0.1875      2.23  0.02644 *
## Year2007          0.3193      0.1847      1.73  0.08497 .
## Year2008          0.4728      0.1841      2.57  0.01071 *
## Year2009          0.3943      0.1804      2.19  0.02965 *
## Year2010          0.4546      0.1830      2.48  0.01352 *
## Year2011          0.4672      0.1792      2.61  0.00956 **
## Year2012          0.4094      0.1782      2.30  0.02227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.127, Adjusted R-squared:  0.0751
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 301 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.397  0.888  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      3.11e-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 1.253 1      1.119
## Year              1.253 16      1.007

```

Residuals from first author



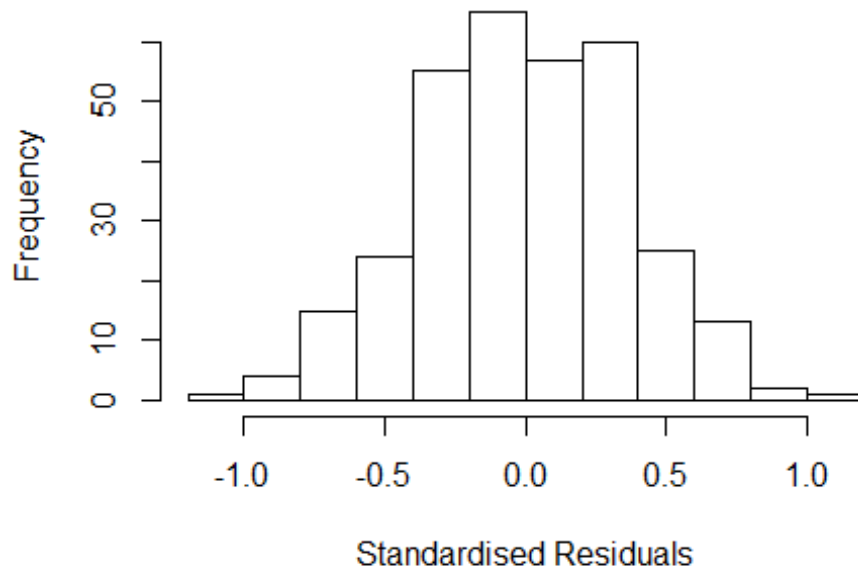
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00459 -0.23708 -0.00361  0.25644  1.01172
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.6842     0.1814   3.77 0.00019 ***
## FirstAuthorFemale1 0.1078     0.0481   2.24 0.02574 *
## Year1997          0.1089     0.2420   0.45 0.65294
## Year1998          0.1534     0.1985   0.77 0.44012
## Year1999          0.1931     0.2053   0.94 0.34782
## Year2000          0.1692     0.1968   0.86 0.39046
## Year2001          0.2058     0.1890   1.09 0.27710
## Year2002          0.1595     0.1888   0.85 0.39874
## Year2003          0.2052     0.1994   1.03 0.30423
## Year2004          0.4123     0.1984   2.08 0.03847 *
## Year2005          0.4788     0.1888   2.54 0.01171 *
## Year2006          0.4270     0.1952   2.19 0.02950 *
```

```

## Year2007          0.3204      0.1949      1.64  0.10129
## Year2008          0.4639      0.1922      2.41  0.01640 *
## Year2009          0.3944      0.1896      2.08  0.03834 *
## Year2010          0.4550      0.1936      2.35  0.01942 *
## Year2011          0.4634      0.1882      2.46  0.01436 *
## Year2012          0.4081      0.1876      2.18  0.03033 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.117, Adjusted R-squared:  0.068
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 294 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.446  0.876  0.950  0.909  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      3.11e-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 1.555 1          1.247
## Year              1.555 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
## -1.01810 -0.24332 0.00186 0.25716 1.09825
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7091 0.1690 4.20 3.6e-05 ***
## LastAuthorFemale1 0.1000 0.0489 2.05 0.042 *
## Year1997 0.1186 0.2350 0.50 0.614
## Year1998 0.1696 0.1931 0.88 0.380
## Year1999 0.2090 0.1993 1.05 0.295
## Year2000 0.1865 0.1890 0.99 0.325
## Year2001 0.2197 0.1841 1.19 0.234
## Year2002 0.1874 0.1818 1.03 0.303
## Year2003 0.2016 0.1929 1.05 0.297
## Year2004 0.4020 0.1864 2.16 0.032 *
## Year2005 0.4682 0.1829 2.56 0.011 *
## Year2006 0.3991 0.1874 2.13 0.034 *
```

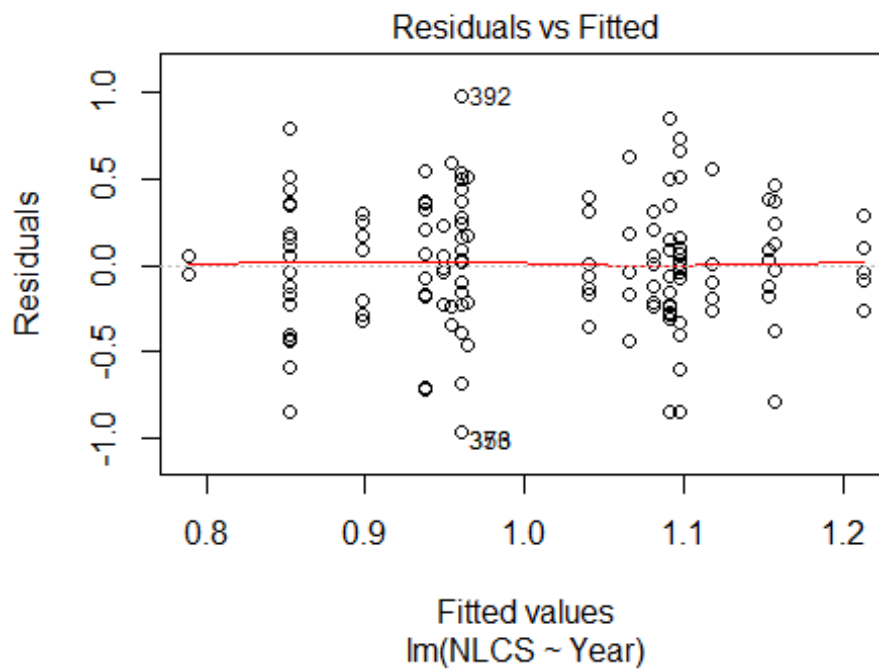
```

## Year2007          0.3060      0.1841      1.66      0.098 .
## Year2008          0.4557      0.1829      2.49      0.013 *
## Year2009          0.3635      0.1775      2.05      0.041 *
## Year2010          0.4455      0.1830      2.43      0.015 *
## Year2011          0.4503      0.1774      2.54      0.012 *
## Year2012          0.3964      0.1778      2.23      0.027 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.116, Adjusted R-squared:  0.0663
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 288 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.370  0.879  0.950  0.906  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      3.11e-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: 322"
## [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
##   11   10   12    5   15   21   11   15   11   11   17   14   14   23   34
## 2011 2012
##   33   39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    4    6    2    5    7    7    7    5    3   11    7    5   14   16
## 2011 2012

```



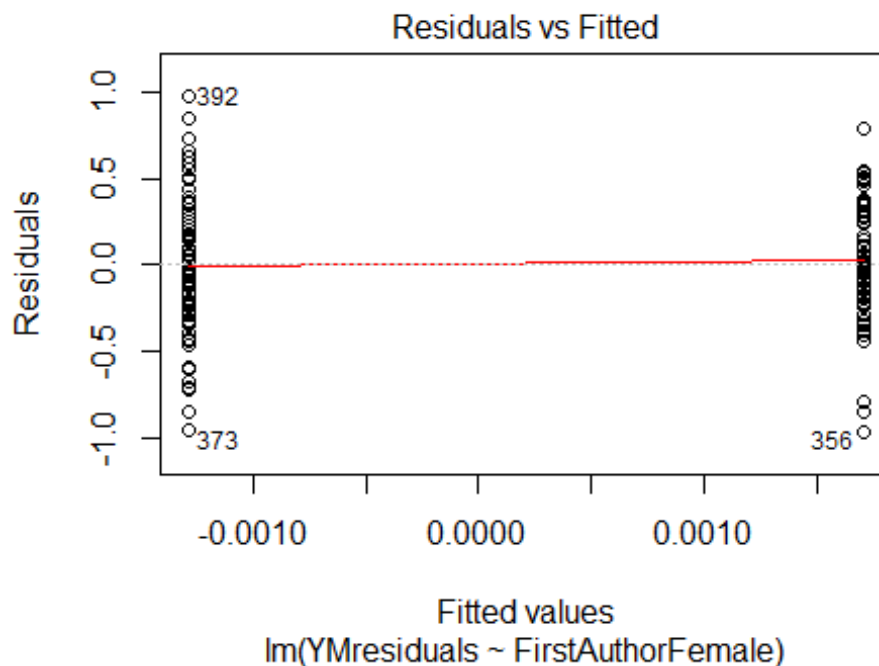
```
## 19 20
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 4 4 2 3 4 7 4 4 2 11 6 4 13 13
## 2011 2012
## 18 19
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```



```
##
## 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: 4.16759866563294"
## [1] "Male first author team size 2018 geometric mean: 5.59508281305979"
## 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 = 25, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.18388952245116"
## [1] "Male last author team size 2018 geometric mean: 5.81562590213848"

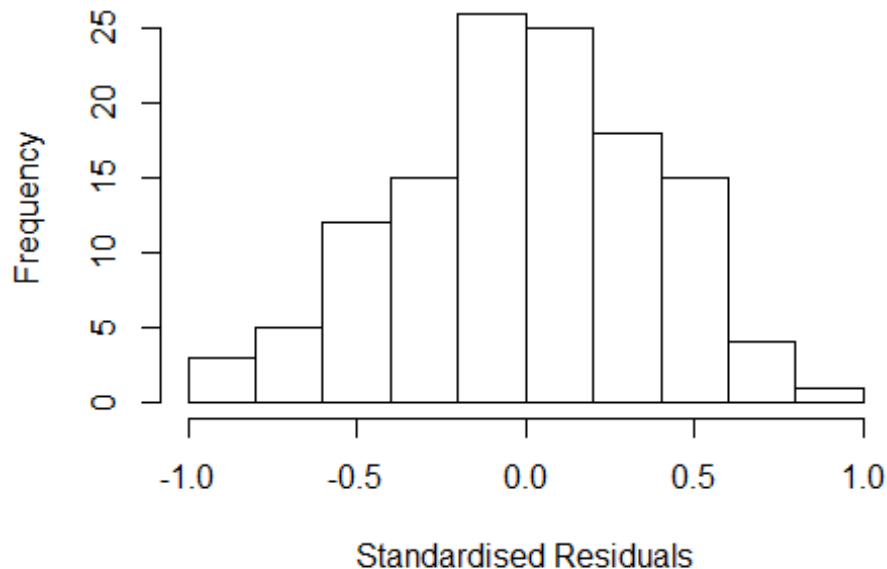
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20, p-value = 0.07
## 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	2.437	1	1.561
## LastAuthorFemale	4.460	1	2.112
## UniqueAuthors	30.143	4	1.531
## Year	135.640	16	1.166

Residuals from first and last author and team size



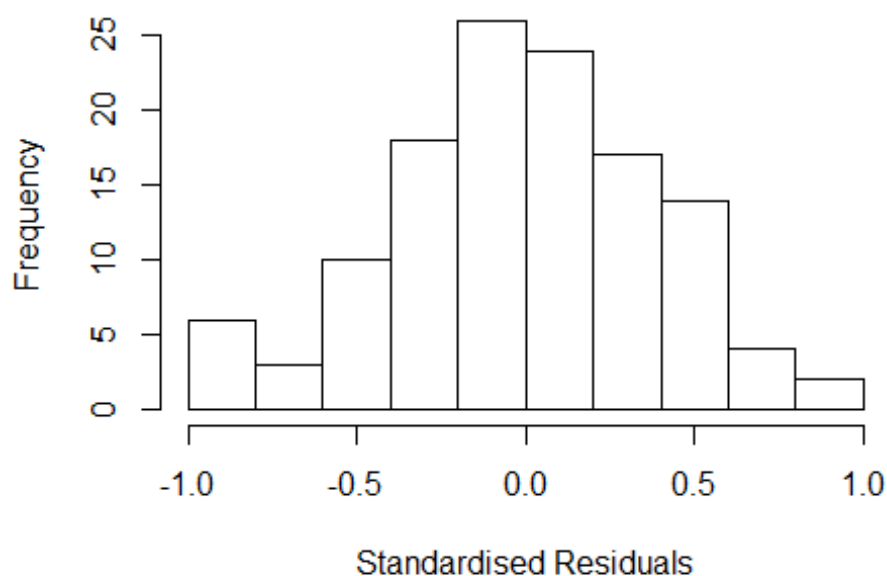
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.92812 -0.24620 0.00244 0.27730 0.83928
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99769 0.15685 6.36 5.9e-09 ***
## FirstAuthorFemale1 -0.00686 0.08472 -0.08 0.936
## LastAuthorFemale1 -0.07154 0.11274 -0.63 0.527
## UniqueAuthors2 0.25375 0.18511 1.37 0.173
## UniqueAuthors3 0.22318 0.16916 1.32 0.190
## UniqueAuthors4 0.16590 0.16412 1.01 0.315
## UniqueAuthors5 0.37091 0.15790 2.35 0.021 *
## Year1997 -0.26495 0.25079 -1.06 0.293
## Year1998 -0.23972 0.26539 -0.90 0.369
## Year1999 -0.43238 0.14214 -3.04 0.003 **
```

```

## Year2000      -0.31008    0.11831   -2.62    0.010 *
## Year2001      -0.17845    0.17512   -1.02    0.311
## Year2002      -0.13728    0.12910   -1.06    0.290
## Year2003      -0.25353    0.24324   -1.04    0.300
## Year2004      -0.19739    0.20396   -0.97    0.335
## Year2005      -0.28968    0.45972   -0.63    0.530
## Year2006      -0.26903    0.16537   -1.63    0.107
## Year2007       0.01473    0.22742    0.06    0.948
## Year2008      -0.07617    0.12512   -0.61    0.544
## Year2009      -0.11305    0.14057   -0.80    0.423
## Year2010      -0.13324    0.17312   -0.77    0.443
## Year2011      -0.31647    0.17072   -1.85    0.067 .
## Year2012      -0.27388    0.18483   -1.48    0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.132, Adjusted R-squared:  -0.0565
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 112 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.592  0.890  0.952  0.921  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      8.06e-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 2.195 1 1.481
## LastAuthorFemale 4.731 1 2.175
## Year 10.325 16 1.076

```

Residuals from first and last author



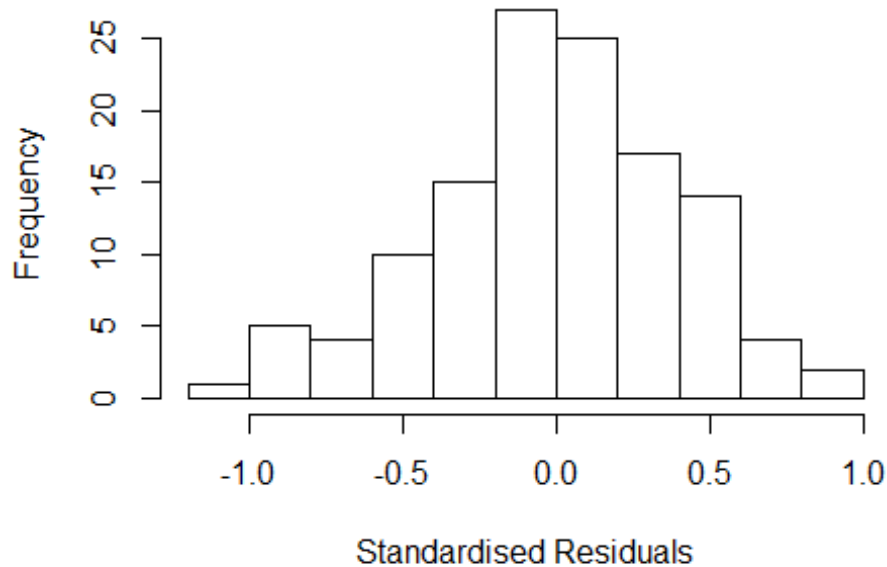
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.98264 -0.22664 -0.00869 0.28494 0.95136
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16127 0.09879 11.75 < 2e-16 ***
## FirstAuthorFemale1 0.02508 0.08344 0.30 0.76436
## LastAuthorFemale1 -0.03898 0.10834 -0.36 0.71969
## Year1997 -0.21074 0.25158 -0.84 0.40412
## Year1998 -0.14568 0.24223 -0.60 0.54886
## Year1999 -0.37277 0.10662 -3.50 0.00069 ***
## Year2000 -0.27335 0.10254 -2.67 0.00889 **
## Year2001 -0.13904 0.15667 -0.89 0.37686
## Year2002 -0.13399 0.13595 -0.99 0.32662
## Year2003 -0.23235 0.17396 -1.34 0.18455
## Year2004 -0.05332 0.20507 -0.26 0.79538
## Year2005 -0.09831 0.49838 -0.20 0.84401
```

```

## Year2006      -0.19662    0.16165   -1.22  0.22659
## Year2007      0.06341    0.22613    0.28  0.77970
## Year2008     -0.00178    0.10173   -0.02  0.98608
## Year2009     -0.03622    0.14565   -0.25  0.80408
## Year2010     -0.06465    0.16385   -0.39  0.69396
## Year2011     -0.27079    0.13132   -2.06  0.04167 *
## Year2012     -0.17863    0.16815   -1.06  0.29053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.064, Adjusted R-squared:  -0.0965
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.497  0.881  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      8.06e-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 2.11 1      1.453
## Year      2.11 16      1.024

```

Residuals from first author



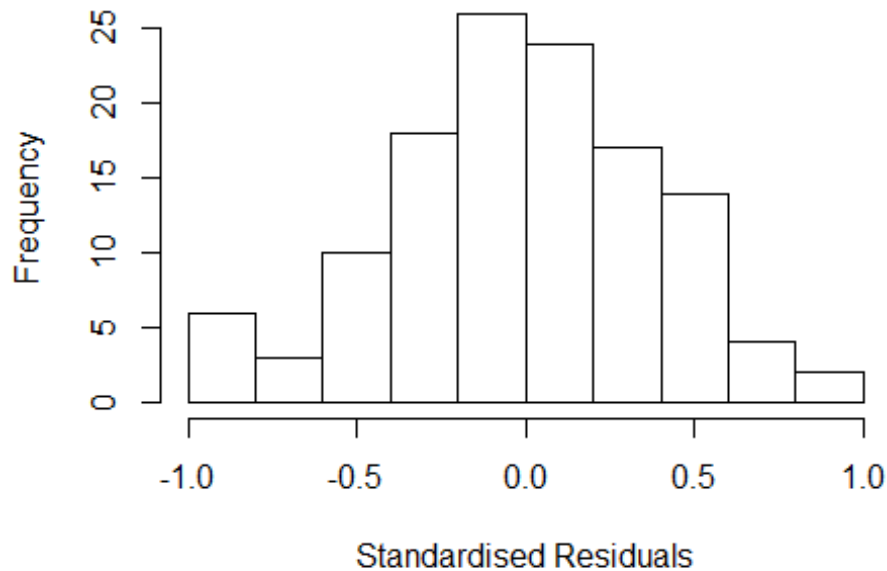
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.001923 -0.218849  0.000567  0.298508  0.955106
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.14179    0.08083   14.13 < 2e-16 ***
## FirstAuthorFemale1  0.02303    0.08194    0.28  0.77922
## Year1997        -0.19592    0.23084   -0.85  0.39795
## Year1998        -0.12237    0.23210   -0.53  0.59914
## Year1999        -0.35329    0.09023   -3.92  0.00016 ***
## Year2000        -0.29219    0.09535   -3.06  0.00277 **
## Year2001        -0.11777    0.14372   -0.82  0.41437
## Year2002        -0.11904    0.13115   -0.91  0.36612
## Year2003        -0.24235    0.16601   -1.46  0.14730
## Year2004        -0.03098    0.19421   -0.16  0.87358
## Year2005        -0.07781    0.47799   -0.16  0.87100
## Year2006        -0.18612    0.15561   -1.20  0.23435
```

```

## Year2007          0.05455    0.22319    0.24  0.80738
## Year2008          -0.00122    0.10422   -0.01  0.99067
## Year2009          -0.03591    0.14689   -0.24  0.80732
## Year2010          -0.05971    0.16101   -0.37  0.71150
## Year2011          -0.26398    0.13086   -2.02  0.04619 *
## Year2012          -0.16290    0.15154   -1.07  0.28485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0615, Adjusted R-squared:  -0.089
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.509  0.886  0.954  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      8.06e-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 4.353 1      2.086
## Year            4.353 16      1.047

```


Residuals from last author



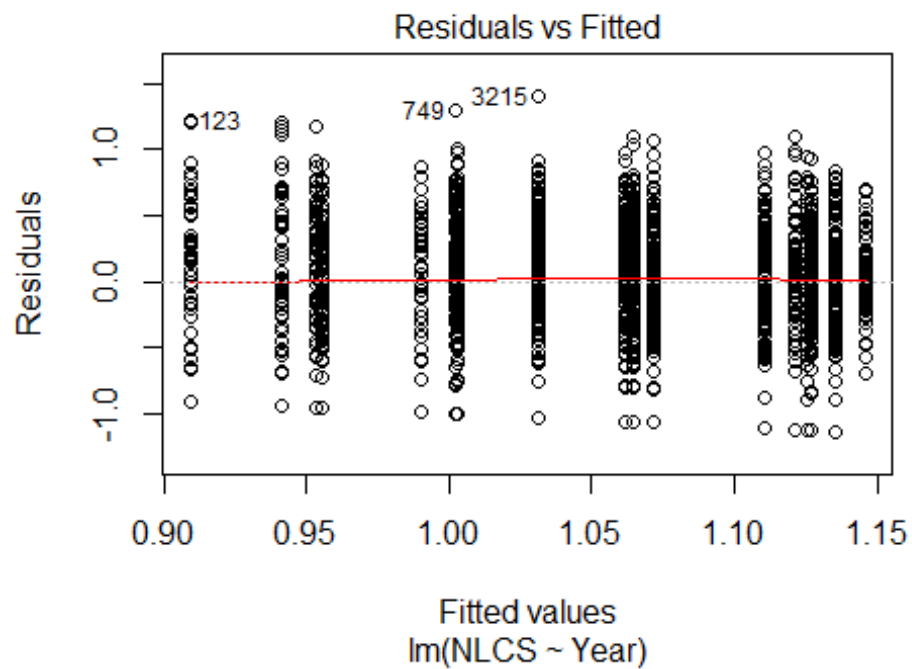
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -9.92e-01 -2.16e-01 -8.51e-05 2.94e-01 9.42e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17028 0.09962 11.75 < 2e-16 ***
## LastAuthorFemale1 -0.04099 0.10723 -0.38 0.70306
## Year1997 -0.20275 0.23877 -0.85 0.39772
## Year1998 -0.13614 0.23132 -0.59 0.55744
## Year1999 -0.38178 0.10738 -3.56 0.00057 ***
## Year2000 -0.27198 0.10449 -2.60 0.01057 *
## Year2001 -0.12899 0.15133 -0.85 0.39594
## Year2002 -0.12725 0.12938 -0.98 0.32759
## Year2003 -0.23347 0.17658 -1.32 0.18897
## Year2004 -0.04514 0.19494 -0.23 0.81734
## Year2005 -0.09478 0.45312 -0.21 0.83472
## Year2006 -0.19918 0.16089 -1.24 0.21846
```

```

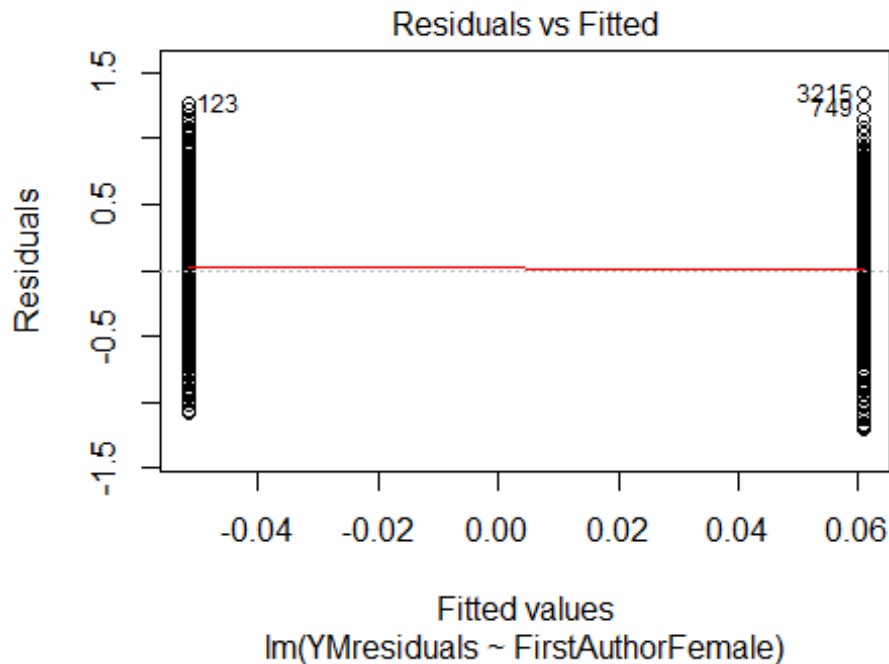
## Year2007      0.06263      0.22164      0.28  0.77807
## Year2008     -0.00979      0.10202     -0.10  0.92377
## Year2009     -0.03929      0.14598     -0.27  0.78833
## Year2010     -0.06066      0.15915     -0.38  0.70385
## Year2011     -0.26966      0.13265     -2.03  0.04456 *
## Year2012     -0.17848      0.16551     -1.08  0.28333
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0621, Adjusted R-squared:  -0.0884
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.536  0.888  0.958  0.916  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      8.06e-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: 124"
## [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
## 104 119 119 126 122 114 108 136 150 167 198 195 240 252 282
## 2011 2012
## 277 278
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 66 70 78 63 54 68 83 100 116 138 132 170 172 184
## 2011 2012

```

```
## 198 195
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 56 59 69 59 47 65 71 87 98 122 118 149 150 170
## 2011 2012
## 184 170
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 47, df = 16, p-value = 6e-05
```

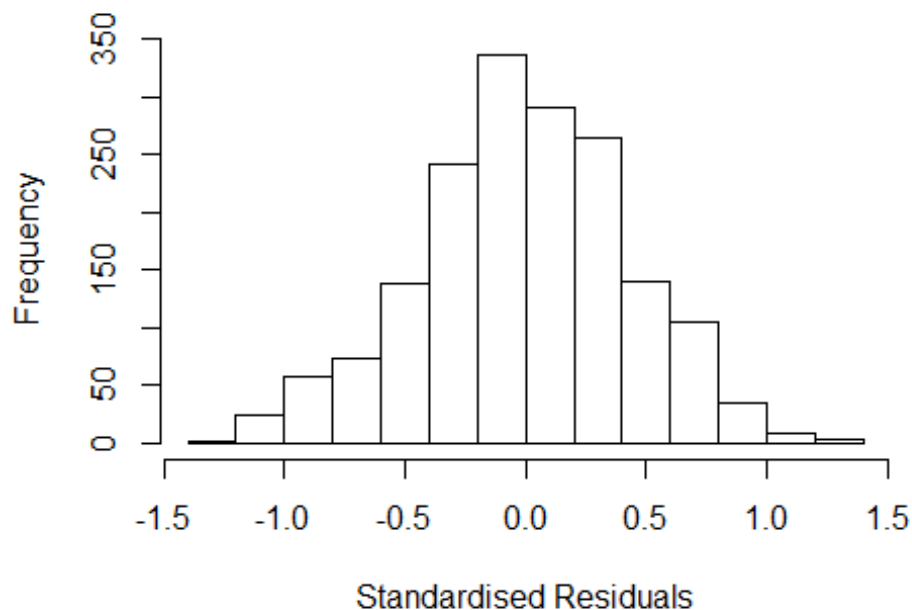


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 6.04003919005485"
## [1] "Male first author team size 2018 geometric mean: 4.45055377742444"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5300, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.25815248478966"
## [1] "Male last author team size 2018 geometric mean: 5.0942447369634"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4000, 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 1.118 1 1.057
## LastAuthorFemale 1.141 1 1.068
## UniqueAuthors 1.365 4 1.040
## Year 1.458 16 1.012
```

Residuals from first and last author and team size



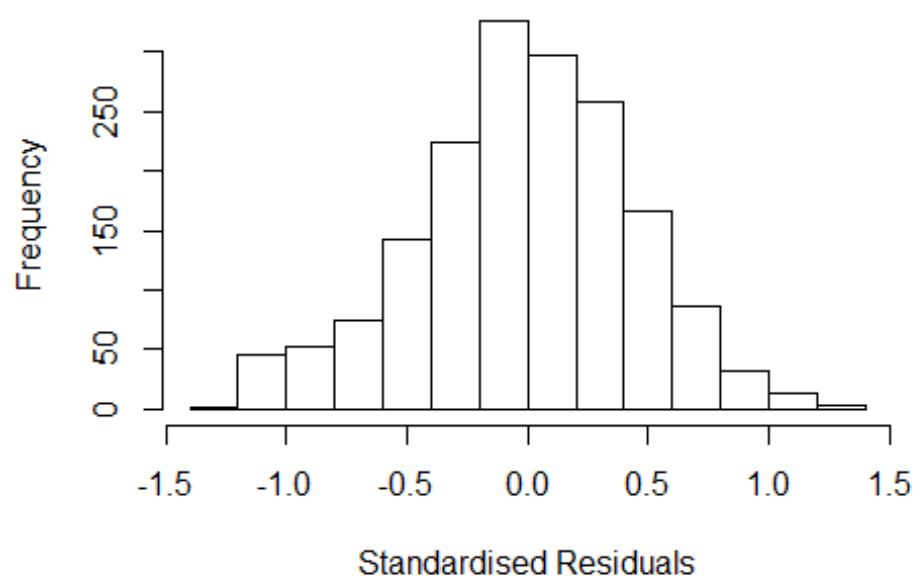
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2301 -0.2688 -0.0068  0.2772  1.3576
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76473    0.08566   8.93 < 2e-16 ***
## FirstAuthorFemale1 0.08054    0.02234   3.61 0.00032 ***
## LastAuthorFemale1 0.03573    0.02456   1.45 0.14591
## UniqueAuthors2    0.19479    0.05956   3.27 0.00110 **
## UniqueAuthors3    0.23338    0.05762   4.05 5.3e-05 ***
## UniqueAuthors4    0.29044    0.05774   5.03 5.4e-07 ***
## UniqueAuthors5    0.38158    0.05334   7.15 1.2e-12 ***
## Year1997         -0.14565    0.10652  -1.37 0.17170
## Year1998         -0.06330    0.12370  -0.51 0.60889
## Year1999         -0.09196    0.09773  -0.94 0.34687
```

```

## Year2000          0.04759      0.09297      0.51  0.60882
## Year2001         -0.05412      0.10258     -0.53  0.59783
## Year2002          0.03528      0.07987      0.44  0.65877
## Year2003          0.02278      0.08194      0.28  0.78108
## Year2004         -0.13684      0.08403     -1.63  0.10362
## Year2005          0.00204      0.07980      0.03  0.97957
## Year2006          0.02627      0.08034      0.33  0.74370
## Year2007         -0.02532      0.07982     -0.32  0.75116
## Year2008         -0.07751      0.07821     -0.99  0.32180
## Year2009         -0.00489      0.07972     -0.06  0.95109
## Year2010         -0.03253      0.07768     -0.42  0.67544
## Year2011         -0.05855      0.07623     -0.77  0.44256
## Year2012          0.04535      0.07723      0.59  0.55714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.0919, Adjusted R-squared:  0.0801
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.861  0.952  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      5.80e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.103 1      1.050
## LastAuthorFemale  1.117 1      1.057
## Year              1.124 16      1.004

```

Residuals from first and last author



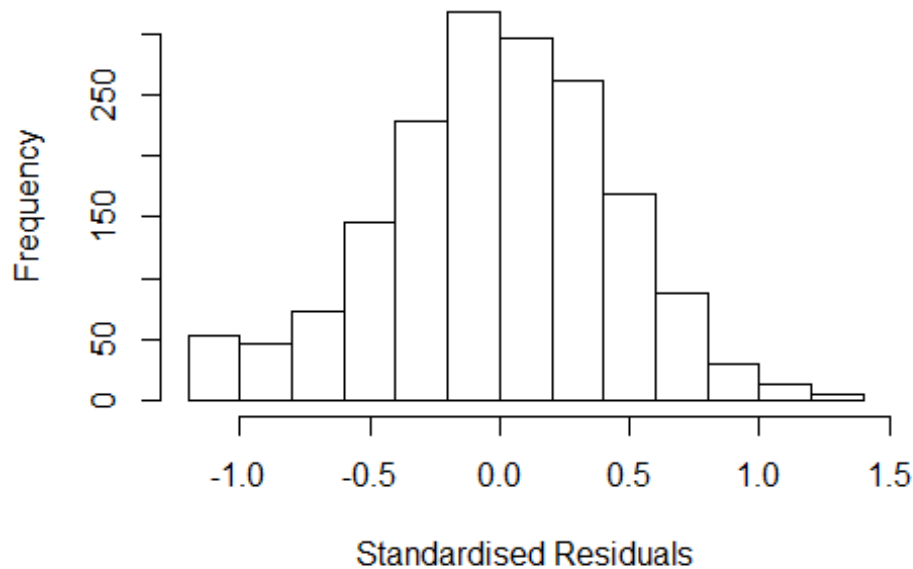
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2116 -0.2869 -0.0049 0.2927 1.3175
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9878 0.0703 14.04 < 2e-16 ***
## FirstAuthorFemale1 0.0981 0.0228 4.30 1.8e-05 ***
## LastAuthorFemale1 0.0177 0.0246 0.72 0.47
## Year1997 -0.1475 0.1099 -1.34 0.18
## Year1998 -0.0557 0.1194 -0.47 0.64
## Year1999 -0.0687 0.0972 -0.71 0.48
## Year2000 0.0813 0.0919 0.88 0.38
## Year2001 -0.0237 0.1109 -0.21 0.83
## Year2002 0.0934 0.0779 1.20 0.23
## Year2003 0.0894 0.0815 1.10 0.27
## Year2004 -0.0840 0.0848 -0.99 0.32
## Year2005 0.0550 0.0792 0.70 0.49
```

```

## Year2006          0.1029      0.0781      1.32      0.19
## Year2007          0.0398      0.0792      0.50      0.62
## Year2008         -0.0270      0.0790     -0.34      0.73
## Year2009          0.0484      0.0783      0.62      0.54
## Year2010          0.0385      0.0778      0.50      0.62
## Year2011          0.0110      0.0758      0.15      0.88
## Year2012          0.1080      0.0764      1.41      0.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0343, Adjusted R-squared:  0.0241
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1584 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.323  0.865   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      5.80e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



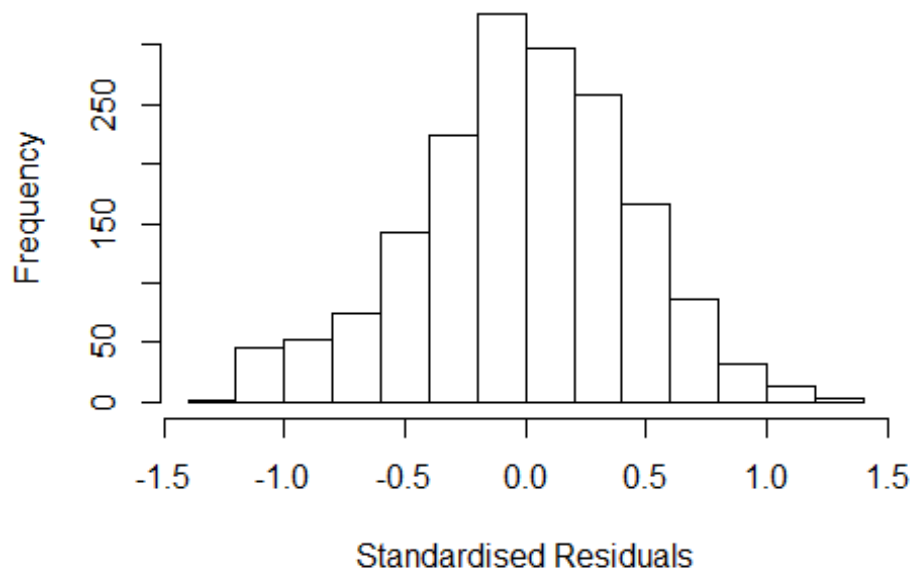
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19984 -0.28493 -0.00104 0.29071 1.32750
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9915 0.0698 14.20 < 2e-16 ***
## FirstAuthorFemale1 0.1017 0.0224 4.55 5.8e-06 ***
## Year1997 -0.1478 0.1098 -1.35 0.18
## Year1998 -0.0566 0.1190 -0.48 0.63
## Year1999 -0.0686 0.0974 -0.70 0.48
## Year2000 0.0816 0.0919 0.89 0.38
## Year2001 -0.0232 0.1107 -0.21 0.83
## Year2002 0.0914 0.0777 1.18 0.24
## Year2003 0.0908 0.0814 1.12 0.26
## Year2004 -0.0851 0.0847 -1.00 0.32
## Year2005 0.0563 0.0792 0.71 0.48
## Year2006 0.1017 0.0780 1.30 0.19
```

```

## Year2007          0.0390      0.0791      0.49      0.62
## Year2008         -0.0280      0.0788     -0.35      0.72
## Year2009          0.0488      0.0782      0.62      0.53
## Year2010          0.0381      0.0777      0.49      0.62
## Year2011          0.0113      0.0757      0.15      0.88
## Year2012          0.1067      0.0762      1.40      0.16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.0243
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1578 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.316  0.864  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      5.80e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



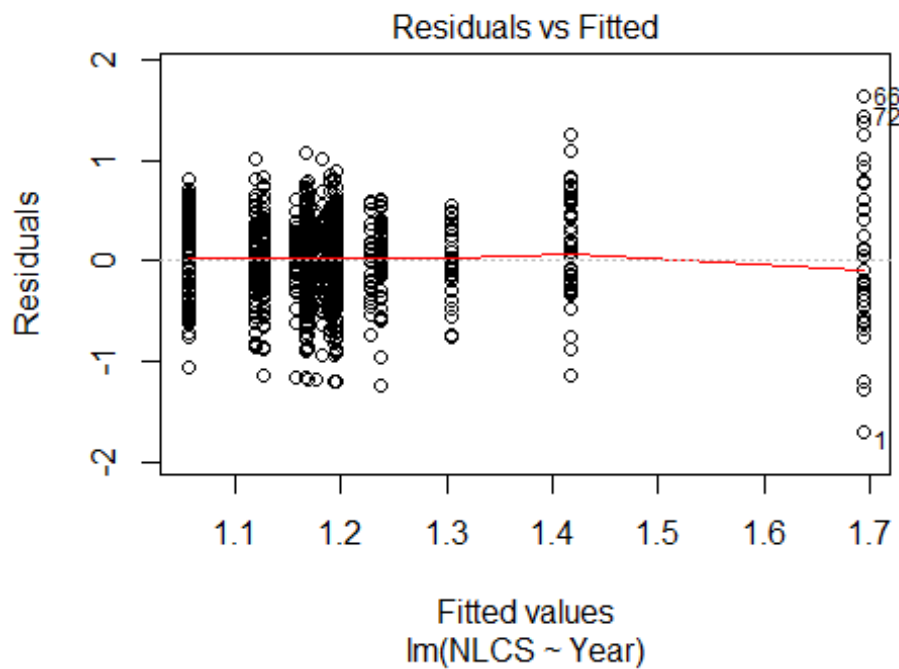
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17763 -0.28439 0.00815 0.28405 1.35645
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0210 0.0718 14.22 <2e-16 ***
## LastAuthorFemale1 0.0419 0.0241 1.74 0.083 .
## Year1997 -0.1390 0.1136 -1.22 0.221
## Year1998 -0.0698 0.1224 -0.57 0.568
## Year1999 -0.0617 0.0991 -0.62 0.534
## Year2000 0.0897 0.0937 0.96 0.338
## Year2001 -0.0274 0.1122 -0.24 0.807
## Year2002 0.1048 0.0805 1.30 0.193
## Year2003 0.0930 0.0836 1.11 0.266
## Year2004 -0.0709 0.0859 -0.83 0.409
## Year2005 0.0715 0.0812 0.88 0.379
## Year2006 0.1098 0.0801 1.37 0.170
```

```

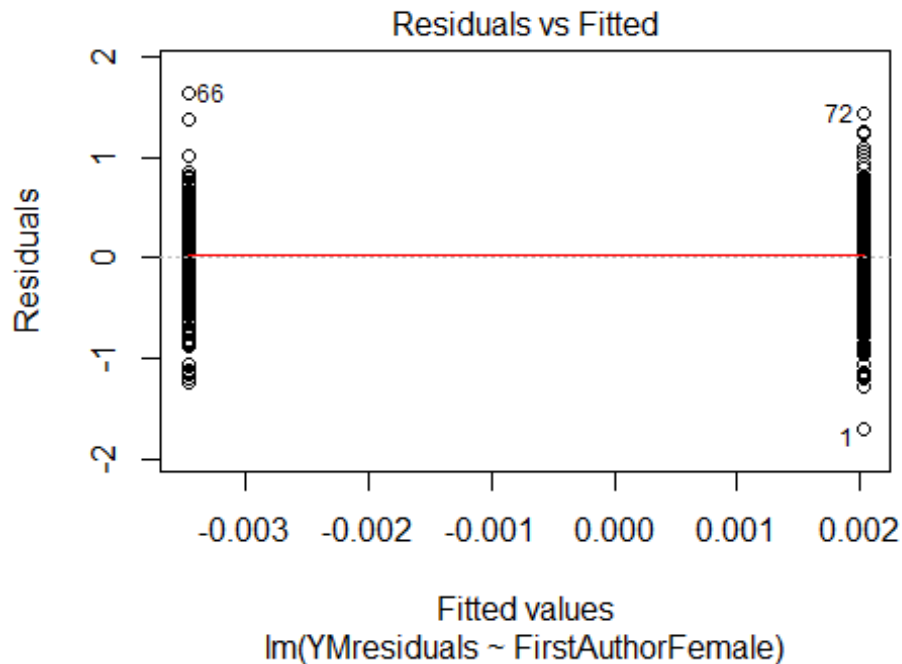
## Year2007          0.0450      0.0813      0.55      0.580
## Year2008         -0.0192      0.0808     -0.24      0.812
## Year2009          0.0538      0.0801      0.67      0.502
## Year2010          0.0549      0.0792      0.69      0.488
## Year2011          0.0127      0.0778      0.16      0.871
## Year2012          0.1147      0.0782      1.47      0.142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0231, Adjusted R-squared:  0.0134
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1576 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.294  0.864  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      5.80e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1723"
## [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
##  102  95   70   60   62   76   73   77   85   84  116  106  114  150  140
## 2011 2012
##  155  138
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   37   52   40   37   28   33   46   52   45   73   84   82   83   98  104
## 2011 2012

```

```
## 116 102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 48 34 36 26 26 37 44 37 59 70 72 77 82 83
## 2011 2012
## 100 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 = 110, df = 16, p-value = 9e-16
```



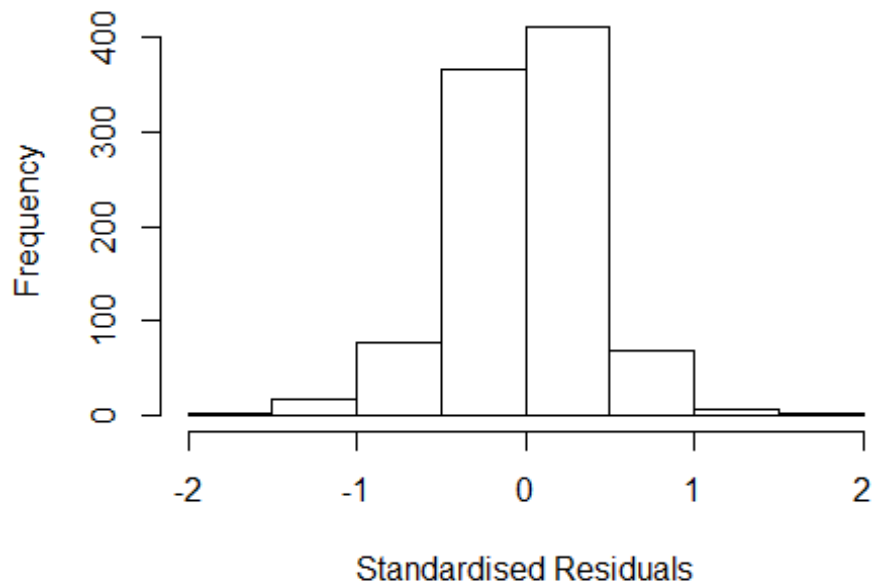
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.032, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 4.64641627328333"
## [1] "Male first author team size 2018 geometric mean: 3.8923044229834"
## 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 = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.09833619452576"
## [1] "Male last author team size 2018 geometric mean: 4.31717634043409"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 480, 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	1.141	1	1.068
LastAuthorFemale	1.183	1	1.088
UniqueAuthors	1.751	4	1.073
Year	2.130	16	1.024

Residuals from first and last author and team size



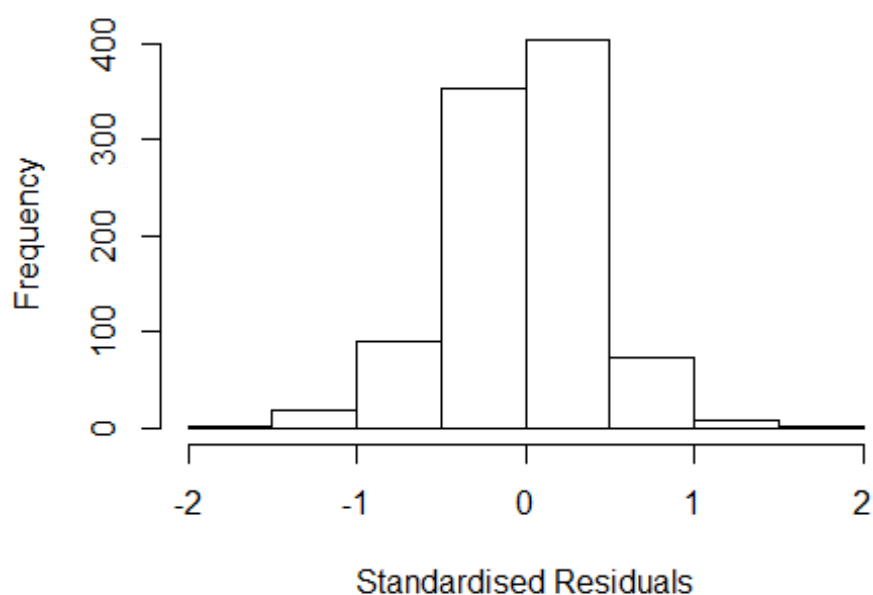
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7061 -0.2478 0.0106 0.2392 1.6446
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.54654 0.17248 8.97 < 2e-16 ***
## FirstAuthorFemale1 -0.02907 0.02723 -1.07 0.28599
## LastAuthorFemale1 0.00385 0.03274 0.12 0.90638
## UniqueAuthors2 0.16007 0.06058 2.64 0.00837 **
## UniqueAuthors3 0.15566 0.06094 2.55 0.01079 *
## UniqueAuthors4 0.27545 0.06383 4.32 1.8e-05 ***
## UniqueAuthors5 0.35106 0.05955 5.90 5.2e-09 ***
## Year1997 -0.24732 0.20262 -1.22 0.22254
## Year1998 -0.57767 0.18201 -3.17 0.00155 **
## Year1999 -0.48882 0.17787 -2.75 0.00611 **
```

```

## Year2000      -0.46211      0.18536      -2.49      0.01284 *
## Year2001      -0.40702      0.18789      -2.17      0.03054 *
## Year2002      -0.51736      0.18601      -2.78      0.00552 **
## Year2003      -0.53117      0.18529      -2.87      0.00424 **
## Year2004      -0.60567      0.17497      -3.46      0.00056 ***
## Year2005      -0.58653      0.17475      -3.36      0.00082 ***
## Year2006      -0.57573      0.17606      -3.27      0.00111 **
## Year2007      -0.63818      0.17635      -3.62      0.00031 ***
## Year2008      -0.54075      0.17788      -3.04      0.00243 **
## Year2009      -0.55393      0.17656      -3.14      0.00176 **
## Year2010      -0.58981      0.17819      -3.31      0.00097 ***
## Year2011      -0.58313      0.17452      -3.34      0.00087 ***
## Year2012      -0.66421      0.17825      -3.73      0.00021 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.363
## Multiple R-squared:  0.135, Adjusted R-squared:  0.114
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## observation 1 is an outlier with |weight| = 0 ( < 0.00011);
## 84 weights are ~= 1. The remaining 866 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0045 0.8630 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           1.05e-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 1.104 1 1.051
## LastAuthorFemale 1.125 1 1.061
## Year 1.215 16 1.006

```


Residuals from first and last author



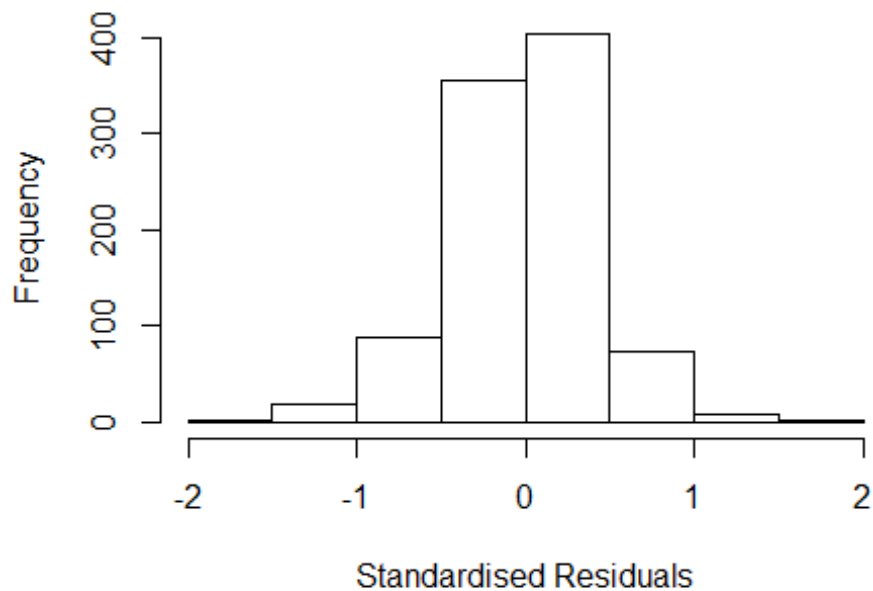
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6718 -0.2642  0.0157  0.2582  1.6592
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.68248    0.18971   8.87  <2e-16 ***
## FirstAuthorFemale1 -0.00506    0.02785  -0.18  0.8557
## LastAuthorFemale1 -0.01064    0.03409  -0.31  0.7551
## Year1997        -0.20945    0.21172  -0.99  0.3228
## Year1998        -0.53128    0.19736  -2.69  0.0072 **
## Year1999        -0.47809    0.19313  -2.48  0.0135 *
## Year2000        -0.43087    0.19934  -2.16  0.0309 *
## Year2001        -0.34836    0.20506  -1.70  0.0897 .
## Year2002        -0.41889    0.20051  -2.09  0.0370 *
## Year2003        -0.48566    0.20174  -2.41  0.0163 *
## Year2004        -0.52926    0.18993  -2.79  0.0054 **
## Year2005        -0.51238    0.19155  -2.67  0.0076 **
```

```

## Year2006          -0.48911    0.19135   -2.56   0.0107 *
## Year2007          -0.56848    0.19259   -2.95   0.0032 **
## Year2008          -0.47032    0.19224   -2.45   0.0146 *
## Year2009          -0.51532    0.19299   -2.67   0.0077 **
## Year2010          -0.49428    0.19346   -2.55   0.0108 *
## Year2011          -0.50499    0.19112   -2.64   0.0084 **
## Year2012          -0.59983    0.19550   -3.07   0.0022 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.0776, Adjusted R-squared:  0.0598
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 871 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0177 0.8650 0.9500 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      1.05e-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 1.094 1      1.046
## Year      1.094 16      1.003

```

Residuals from first author



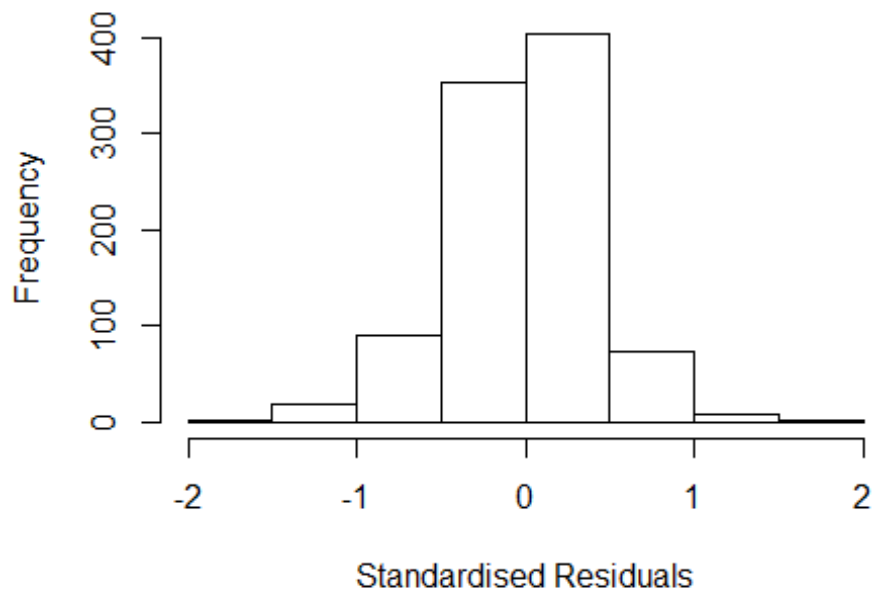
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.679 -0.262 0.013 0.262 1.654
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.67870 0.18641 9.01 <2e-16 ***
## FirstAuthorFemale1 -0.00643 0.02776 -0.23 0.8170
## Year1997 -0.20927 0.21060 -0.99 0.3206
## Year1998 -0.52862 0.19489 -2.71 0.0068 **
## Year1999 -0.47457 0.19004 -2.50 0.0127 *
## Year2000 -0.42833 0.19725 -2.17 0.0301 *
## Year2001 -0.34581 0.20245 -1.71 0.0879 .
## Year2002 -0.41571 0.19750 -2.10 0.0356 *
## Year2003 -0.48410 0.20012 -2.42 0.0158 *
## Year2004 -0.52800 0.18835 -2.80 0.0052 **
## Year2005 -0.51101 0.18994 -2.69 0.0073 **
## Year2006 -0.48785 0.18975 -2.57 0.0103 *
```

```

## Year2007          -0.56615    0.19040   -2.97    0.0030 **
## Year2008          -0.46870    0.19040   -2.46    0.0140 *
## Year2009          -0.51316    0.19059   -2.69    0.0072 **
## Year2010          -0.49271    0.19157   -2.57    0.0103 *
## Year2011          -0.50367    0.18942   -2.66    0.0080 **
## Year2012          -0.59856    0.19395   -3.09    0.0021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.0772, Adjusted R-squared:  0.0604
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 872 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0168 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      1.05e-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 1.116 1          1.056
## Year            1.116 16          1.003

```

Residuals from last author



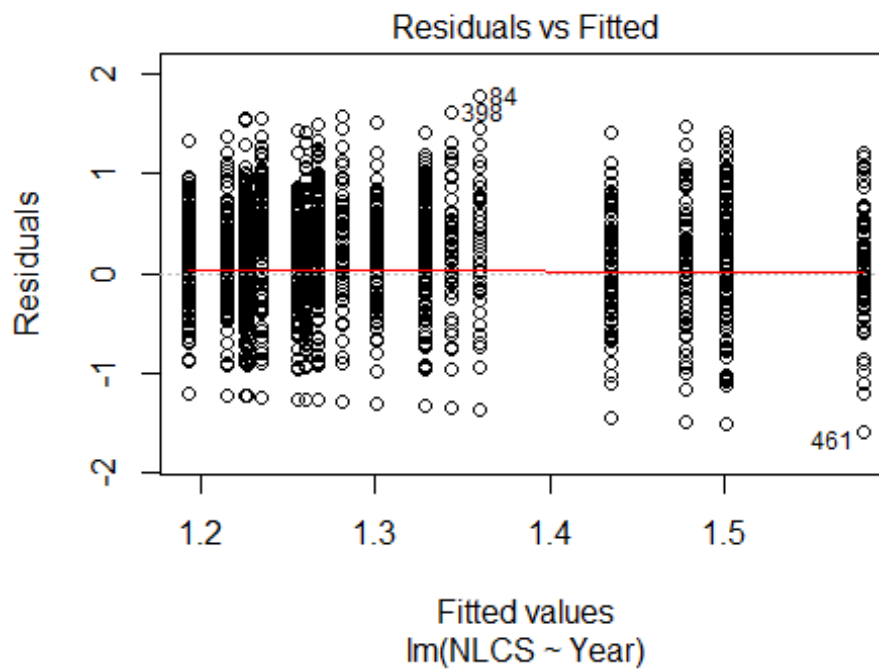
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6684 -0.2648 0.0153 0.2596 1.6576
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6798 0.1890 8.89 <2e-16 ***
## LastAuthorFemale1 -0.0114 0.0340 -0.34 0.7369
## Year1997 -0.2082 0.2118 -0.98 0.3260
## Year1998 -0.5303 0.1975 -2.69 0.0074 **
## Year1999 -0.4774 0.1934 -2.47 0.0138 *
## Year2000 -0.4285 0.1987 -2.16 0.0313 *
## Year2001 -0.3460 0.2046 -1.69 0.0912 .
## Year2002 -0.4182 0.2007 -2.08 0.0375 *
## Year2003 -0.4844 0.2017 -2.40 0.0165 *
## Year2004 -0.5288 0.1902 -2.78 0.0055 **
## Year2005 -0.5117 0.1918 -2.67 0.0078 **
## Year2006 -0.4883 0.1916 -2.55 0.0110 *
```

```

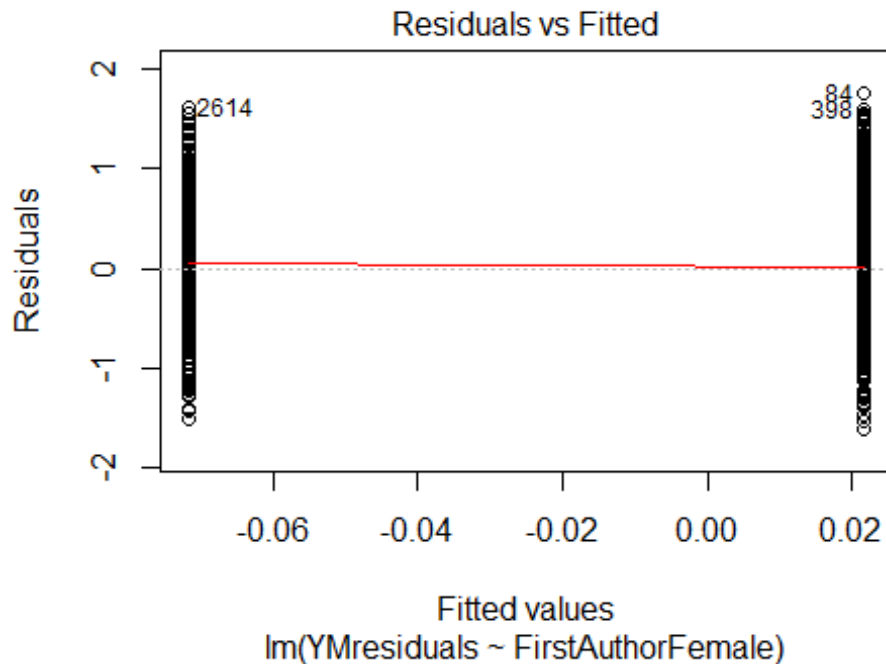
## Year2007          -0.5678      0.1928    -2.94    0.0033 **
## Year2008          -0.4689      0.1922    -2.44    0.0149 *
## Year2009          -0.5144      0.1932    -2.66    0.0079 **
## Year2010          -0.4935      0.1937    -2.55    0.0110 *
## Year2011          -0.5041      0.1913    -2.63    0.0086 **
## Year2012          -0.5987      0.1957    -3.06    0.0023 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.0775, Adjusted R-squared:  0.0607
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 875 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.018  0.866  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      1.05e-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: 951"
## [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
## 122 138 159 157 166 192 198 175 143 188 273 274 288 254 263
## 2011 2012
## 316 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 84 81 83 95 135 161 145 118 157 234 237 249 218 223
## 2011 2012

```

```
## 269 220
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 58 80 79 79 87 125 142 132 106 152 218 224 220 197 208
## 2011 2012
## 251 200
## [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 = 3e-16
```

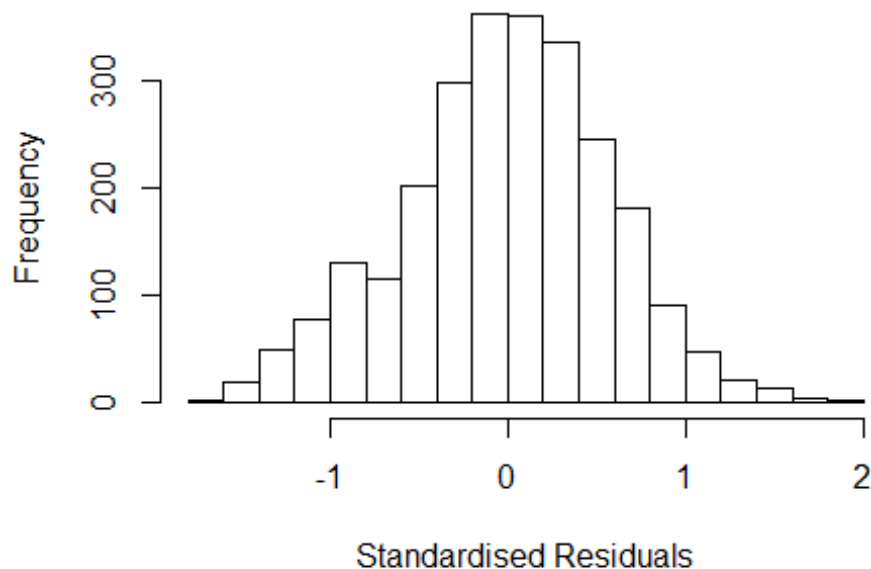


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 28, df = 1, p-value = 1e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 3.46710670774929"
## [1] "Male first author team size 2018 geometric mean: 3.27420618812341"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8300, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.39448497447994"
## [1] "Male last author team size 2018 geometric mean: 3.51441897809227"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8800, 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 1.189 1      1.090
## LastAuthorFemale  1.180 1      1.086
## UniqueAuthors    1.341 4      1.037
## Year             1.379 16      1.010
```


Residuals from first and last author and team size



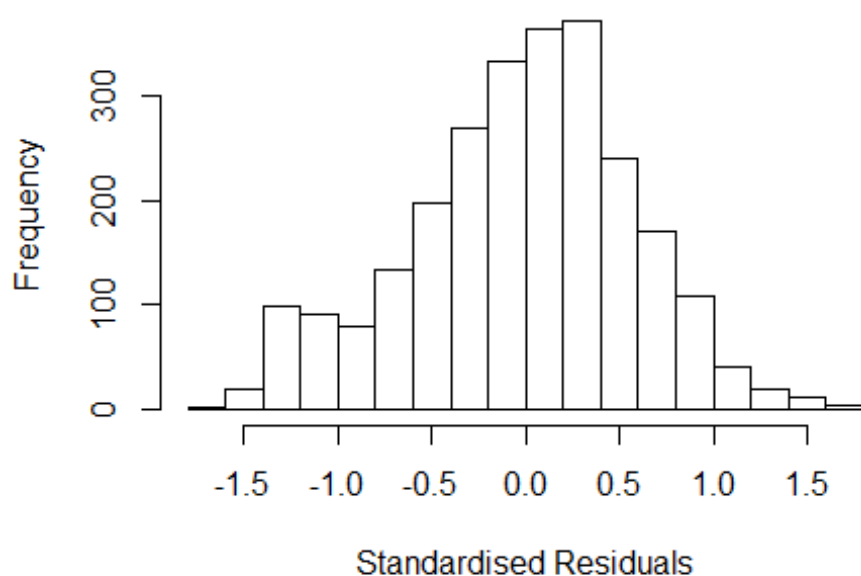
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7494 -0.3703 0.0108 0.3718 1.8677
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13890 0.11109 10.25 < 2e-16 ***
## FirstAuthorFemale1 0.00849 0.03296 0.26 0.7968
## LastAuthorFemale1 0.11892 0.03021 3.94 8.5e-05 ***
## UniqueAuthors2 0.25808 0.03633 7.10 1.6e-12 ***
## UniqueAuthors3 0.36422 0.03716 9.80 < 2e-16 ***
## UniqueAuthors4 0.43060 0.04145 10.39 < 2e-16 ***
## UniqueAuthors5 0.46507 0.04195 11.09 < 2e-16 ***
## Year1997 -0.05152 0.13753 -0.37 0.7080
## Year1998 0.01324 0.13970 0.09 0.9245
## Year1999 0.23354 0.12493 1.87 0.0617 .
```

```

## Year2000      0.00852    0.12761    0.07    0.9468
## Year2001     -0.04859    0.11826   -0.41    0.6812
## Year2002     -0.23468    0.11822   -1.99    0.0472 *
## Year2003     -0.24413    0.11797   -2.07    0.0386 *
## Year2004     -0.19781    0.12432   -1.59    0.1117
## Year2005     -0.26775    0.11471   -2.33    0.0197 *
## Year2006     -0.25741    0.11317   -2.27    0.0230 *
## Year2007     -0.31007    0.11452   -2.71    0.0068 **
## Year2008     -0.30004    0.11373   -2.64    0.0084 **
## Year2009     -0.23010    0.12031   -1.91    0.0559 .
## Year2010     -0.15982    0.11730   -1.36    0.1732
## Year2011     -0.28092    0.12251   -2.29    0.0219 *
## Year2012     -0.03105    0.11707   -0.27    0.7909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.114, Adjusted R-squared:  0.107
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 217 weights are ~= 1. The remaining 2341 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.223  0.863  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      3.91e-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 1.177 1      1.085
## LastAuthorFemale  1.155 1      1.075
## Year              1.092 16      1.003

```

Residuals from first and last author



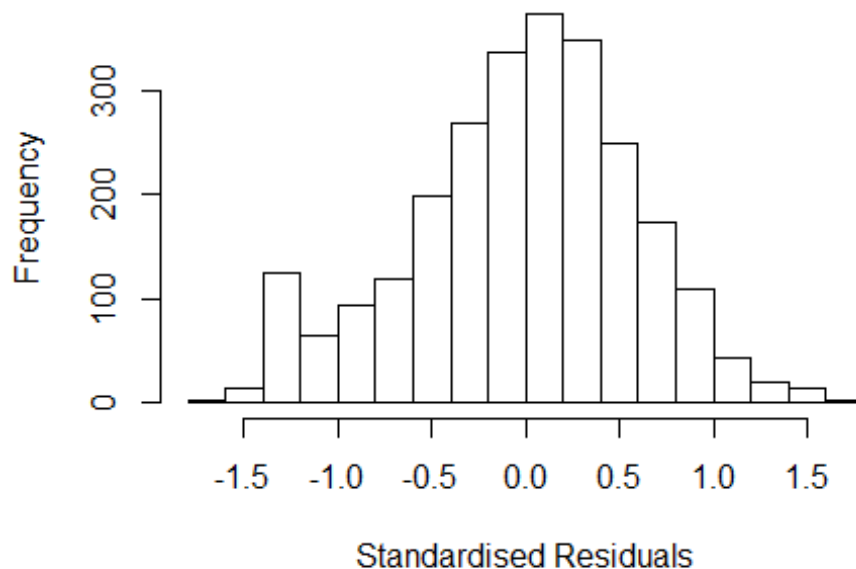
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6549 -0.3773  0.0344  0.3753  1.7114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2662     0.1149   11.02 < 2e-16 ***
## FirstAuthorFemale1  0.0496     0.0353    1.40  0.16074
## LastAuthorFemale1  0.1069     0.0314    3.40  0.00068 ***
## Year1997         -0.0654     0.1431   -0.46  0.64738
## Year1998         -0.0125     0.1407   -0.09  0.92896
## Year1999          0.2322     0.1287    1.80  0.07139 .
## Year2000          0.0613     0.1308    0.47  0.63937
## Year2001          0.0471     0.1236    0.38  0.70335
## Year2002         -0.1495     0.1231   -1.21  0.22492
## Year2003         -0.1228     0.1230   -1.00  0.31803
## Year2004         -0.0896     0.1287   -0.70  0.48659
## Year2005         -0.1382     0.1194   -1.16  0.24723
```

```

## Year2006          -0.1277      0.1179   -1.08   0.27877
## Year2007          -0.1686      0.1194   -1.41   0.15812
## Year2008          -0.1682      0.1200   -1.40   0.16127
## Year2009          -0.0980      0.1264   -0.78   0.43823
## Year2010          -0.0198      0.1226   -0.16   0.87171
## Year2011          -0.1161      0.1279   -0.91   0.36411
## Year2012           0.1534      0.1219    1.26   0.20849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0404, Adjusted R-squared:  0.0336
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 196 weights are ~= 1. The remaining 2362 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.328  0.858   0.951   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      3.91e-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 1.076 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
## -1.6311 -0.3859 0.0309 0.3820 1.7335
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30466 0.11527 11.32 <2e-16 ***
## FirstAuthorFemale1 0.09583 0.03445 2.78 0.0055 **
## Year1997 -0.07189 0.14387 -0.50 0.6174
## Year1998 -0.00591 0.14094 -0.04 0.9666
## Year1999 0.23060 0.12963 1.78 0.0754 .
## Year2000 0.06035 0.13144 0.46 0.6462
## Year2001 0.05706 0.12422 0.46 0.6461
## Year2002 -0.14108 0.12344 -1.14 0.2532
## Year2003 -0.11923 0.12357 -0.96 0.3347
## Year2004 -0.08198 0.12905 -0.64 0.5253
## Year2005 -0.13508 0.11998 -1.13 0.2603
## Year2006 -0.12656 0.11850 -1.07 0.2856
```

```

## Year2007          -0.16199      0.12005    -1.35    0.1773
## Year2008          -0.16522      0.12068    -1.37    0.1711
## Year2009          -0.09347      0.12723    -0.73    0.4626
## Year2010          -0.01456      0.12326    -0.12    0.9060
## Year2011          -0.11552      0.12861    -0.90    0.3692
## Year2012           0.15937      0.12238     1.30    0.1929
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0349, Adjusted R-squared:  0.0284
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 212 weights are ~= 1. The remaining 2346 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.318  0.856  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      3.91e-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 1.054 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
## -1.6479 -0.3743  0.0388  0.3713  1.7243

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28449   0.11321  11.35 < 2e-16 ***
## LastAuthorFemale1 0.12520   0.03029   4.13 3.7e-05 ***
## Year1997       -0.05968   0.14199  -0.42  0.674
## Year1998       -0.00751   0.13936  -0.05  0.957
## Year1999        0.23825   0.12755   1.87  0.062 .
## Year2000        0.06913   0.12933   0.53  0.593
## Year2001        0.05286   0.12240   0.43  0.666
## Year2002       -0.14174   0.12193  -1.16  0.245
## Year2003       -0.11630   0.12171  -0.96  0.339
## Year2004       -0.08129   0.12715  -0.64  0.523
## Year2005       -0.13185   0.11820  -1.12  0.265
## Year2006       -0.11989   0.11649  -1.03  0.304
## Year2007       -0.15989   0.11793  -1.36  0.175
## Year2008       -0.16104   0.11870  -1.36  0.175
## Year2009       -0.08918   0.12502  -0.71  0.476
## Year2010       -0.01251   0.12124  -0.10  0.918
## Year2011       -0.10631   0.12636  -0.84  0.400
## Year2012        0.15941   0.12071   1.32  0.187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0393, Adjusted R-squared:  0.0329
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.321 0.860  0.951  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           3.91e-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: 2558"
## [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]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    4    3    1    4    6    3    6    3    4    6    2    8   14   12    8
## 2012
##    9
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    0    0    0    2    2    3    1    1    2    0    7   12    5    5
## 2012
##    4
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    0    0    0    2    2    2    1    1    1    0    6   10    4    3
## 2012
##    4
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.79606471179477"
## [1] "Male first author team size 2018 geometric mean: 1"

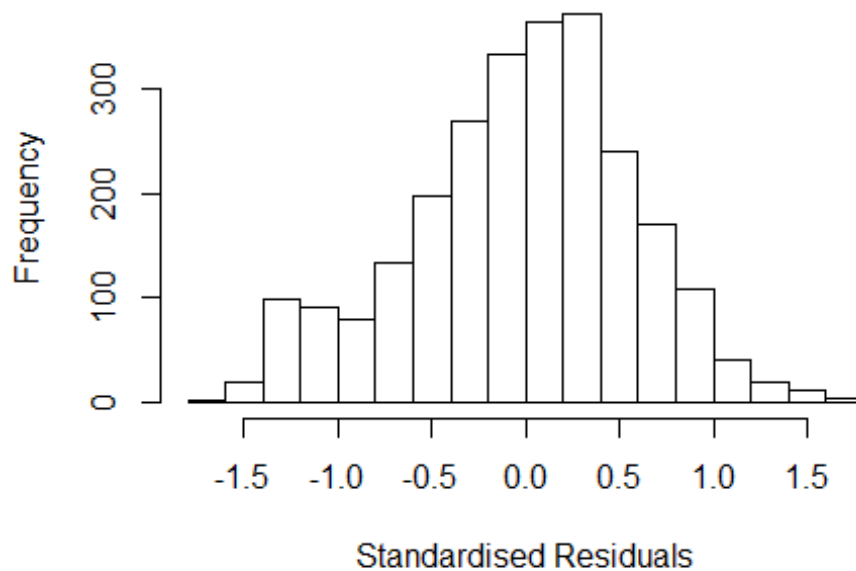
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.22257477860518"
## [1] "Male last author team size 2018 geometric mean: 3.65076273456752"

## 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 = 26, 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 8.569e+14 1      2.927e+07
## LastAuthorFemale 1.515e+15 1      3.893e+07
## UniqueAuthors    5.574e+44 4      3.920e+05
## Year              6.766e+44 11     1.091e+02

## [1] "List of 0 outliers with 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
## -1.61e+00 -1.78e-01 -6.66e-16  1.25e-01  5.34e-01
##
## Coefficients:
##               Estimate Std. Error   t value Pr(>|t|)
## (Intercept)    2.61e+00   2.22e-01   1.18e+01  3.7e-10 ***
## FirstAuthorFemale1 -5.32e-01   1.52e-01  -3.49e+00  0.00243 **
## LastAuthorFemale1  -2.06e-01   1.93e-01  -1.07e+00  0.29932
## UniqueAuthors2     7.47e-01   1.31e-01   5.71e+00  1.7e-05 ***
## UniqueAuthors3     5.04e-01   1.61e-01   3.13e+00  0.00551 **
## UniqueAuthors4     1.70e-01   2.37e-01   7.20e-01  0.48127
## UniqueAuthors5     8.56e-01   2.22e-01   3.86e+00  0.00105 **
## Year2001          -1.87e+00   6.56e-02  -2.86e+01  < 2e-16 ***
## Year2002          -8.85e-01   1.97e-01  -4.49e+00  0.00025 ***
## Year2003          -9.63e-01   2.08e-01  -4.62e+00  0.00019 ***
## Year2004          -4.55e-01   1.53e-08  -2.97e+07  < 2e-16 ***
## Year2005          -1.64e+00   1.31e-01  -1.25e+01  1.3e-10 ***
## Year2006          -1.31e+00   2.22e-01  -5.92e+00  1.1e-05 ***
## Year2008          -1.83e+00   3.11e-01  -5.88e+00  1.2e-05 ***
## Year2009          -1.65e+00   2.40e-01  -6.85e+00  1.5e-06 ***
## Year2010          -1.57e+00   2.55e-01  -6.15e+00  6.5e-06 ***
## Year2011          -1.45e+00   8.59e-02  -1.69e+01  6.9e-13 ***
## Year2012          -1.95e+00   2.46e-01  -7.92e+00  2.0e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.679, Adjusted R-squared:  0.391
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 28 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0323 0.9280 0.9690 0.9160 0.9910 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           2.70e-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"

## 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

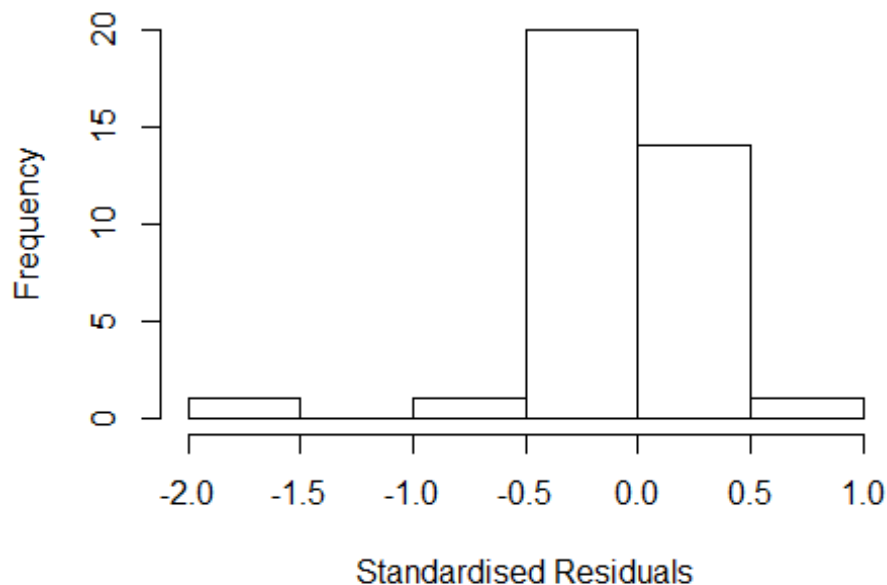
## Warning in lmrob.S(x, y, control = control, mf = mf): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps

## [1] "Regression 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

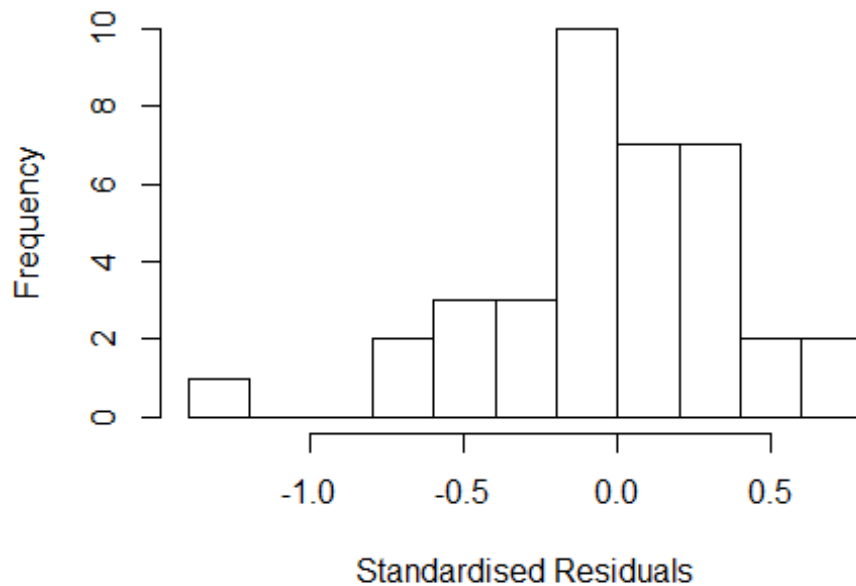
## Warning in cov2cor(v): diag(.) had 0 or NA entries; non-finite result is
## doubtful
```

Residuals from first and last author and team size



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1          NaN
## Year            NaN 11          NaN
```

Residuals from first author



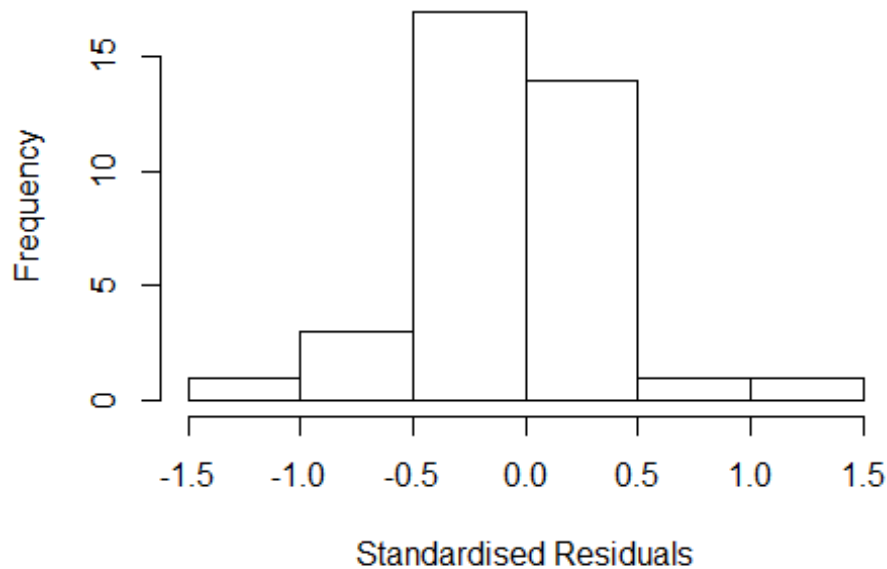
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22e+00 -1.73e-01 -2.22e-16 2.71e-01 6.99e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.09e+00 1.47e-01 1.42e+01 3.7e-13 ***
## FirstAuthorFemale1 -2.21e-01 1.47e-01 -1.50e+00 0.1470
## Year2001 -1.50e+00 2.89e-01 -5.19e+00 2.6e-05 ***
## Year2002 -3.74e-01 1.32e-01 -2.82e+00 0.0094 **
## Year2003 -7.11e-01 4.37e-01 -1.63e+00 0.1171
## Year2004 -4.55e-01 1.71e-08 -2.67e+07 < 2e-16 ***
## Year2005 -8.88e-01 0.00e+00 -Inf < 2e-16 ***
## Year2006 -4.56e-01 0.00e+00 -Inf < 2e-16 ***
## Year2008 -1.08e+00 1.78e-01 -6.08e+00 2.8e-06 ***
## Year2009 -8.71e-01 1.75e-01 -4.98e+00 4.3e-05 ***
## Year2010 -7.84e-01 5.29e-02 -1.48e+01 1.4e-13 ***
## Year2011 -9.53e-01 2.71e-01 -3.52e+00 0.0017 **
```

```

## Year2012          -1.07e+00   1.76e-01 -6.11e+00   2.6e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.367, Adjusted R-squared:  0.0503
## Convergence in 9 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 31 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.540  0.925  0.969  0.938  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          2.70e-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 22.84 1          4.779
## Year            22.84 11          1.153

```

Residuals from last author



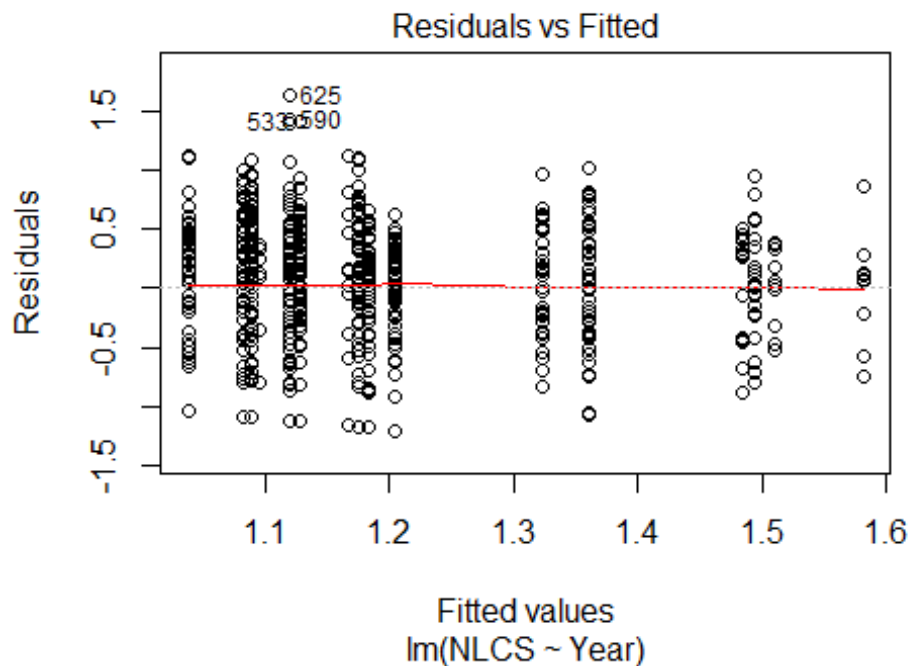
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.02e+00 -2.18e-01 -2.22e-16 2.53e-01 8.98e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.05e+00 1.51e-01 1.36e+01 8.6e-13 ***
## LastAuthorFemale1 -1.86e-01 1.51e-01 -1.23e+00 0.2301
## Year2001 -1.50e+00 2.88e-01 -5.21e+00 2.5e-05 ***
## Year2002 -3.56e-01 1.44e-01 -2.47e+00 0.0211 *
## Year2003 -7.11e-01 4.34e-01 -1.64e+00 0.1143
## Year2004 -4.55e-01 1.93e-08 -2.36e+07 < 2e-16 ***
## Year2005 -8.88e-01 2.09e-08 -4.26e+07 < 2e-16 ***
## Year2006 -4.56e-01 1.41e-08 -3.23e+07 < 2e-16 ***
## Year2008 -1.14e+00 2.15e-01 -5.32e+00 1.9e-05 ***
## Year2009 -8.49e-01 2.21e-01 -3.83e+00 0.0008 ***
## Year2010 -8.68e-01 1.28e-01 -6.78e+00 5.2e-07 ***
## Year2011 -9.41e-01 2.77e-01 -3.40e+00 0.0024 **
```

```

## Year2012          -1.06e+00   1.75e-01 -6.04e+00   3.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.352, Adjusted R-squared:  0.028
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 30 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.673  0.916  0.960  0.938  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           2.70e-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:  37"
## [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
##    21   26   16   21   25   40   33   43   61   58   62   71   93   88   89
## 2011 2012
##    69   72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    12   10    8    9    7   14   26   33   53   49   57   61   77   68   75
## 2011 2012
##    60   57
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    11    9    7    7    6   14   24   27   49   46   56   58   69   66   66
## 2011 2012

```

```
## 49 48
## [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 = 0.17, df = 1, p-value = 0.7

## [1] "Female first author team size 2018 geometric mean: 3.47718853241163"
## [1] "Male first author team size 2018 geometric mean: 4.69605685190299"

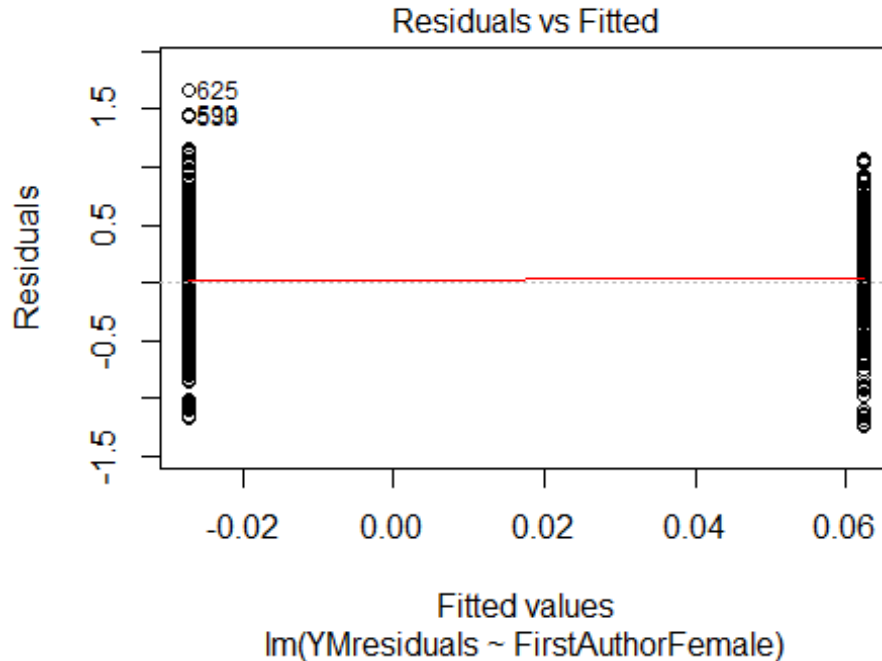
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
```



```
##
## [1] "Female last author team size 2018 geometric mean: 2.55340815750941"
## [1] "Male last author team size 2018 geometric mean: 5.32851385582871"

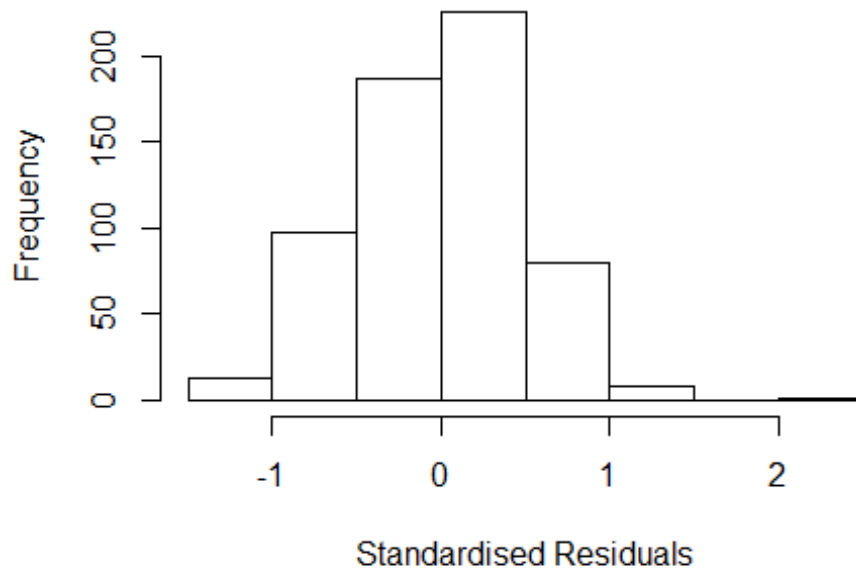
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 7e-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	1.215	1	1.102
LastAuthorFemale	1.382	1	1.175
UniqueAuthors	1.814	4	1.077
Year	1.729	16	1.017

Residuals from first and last author and team size



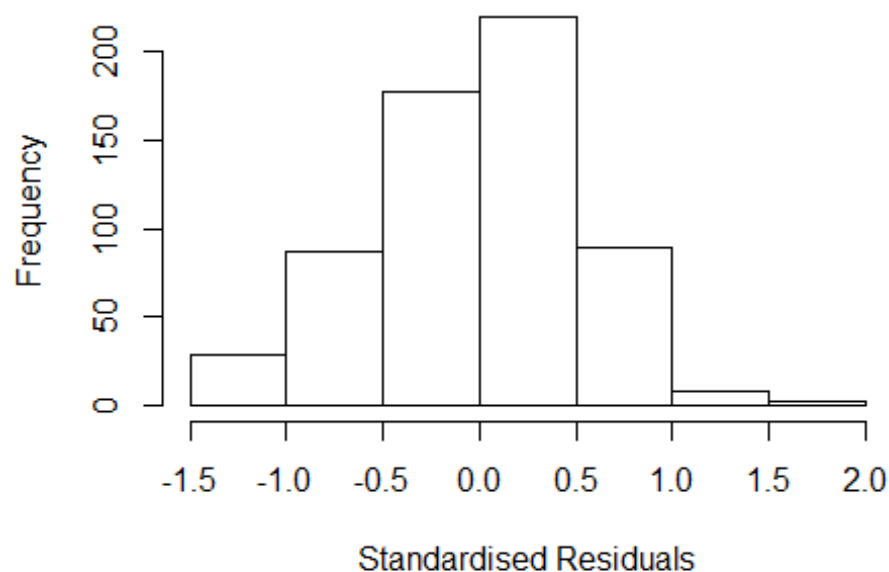
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3430 -0.3148 0.0152 0.3375 2.0697
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8364 0.2186 3.83 0.00014 ***
## FirstAuthorFemale1 -0.0470 0.0471 -1.00 0.31888
## LastAuthorFemale1 -0.0669 0.0467 -1.43 0.15216
## UniqueAuthors2 0.1723 0.0882 1.95 0.05127 .
## UniqueAuthors3 0.3840 0.0953 4.03 6.3e-05 ***
## UniqueAuthors4 0.4836 0.0960 5.04 6.3e-07 ***
## UniqueAuthors5 0.6429 0.0851 7.55 1.6e-13 ***
## Year1997 0.2243 0.2275 0.99 0.32446
## Year1998 0.0824 0.2668 0.31 0.75763
## Year1999 0.3681 0.2659 1.38 0.16683
```

```

## Year2000          -0.0680      0.2437    -0.28  0.78025
## Year2001           0.1906      0.2207      0.86  0.38794
## Year2002           0.2743      0.2213      1.24  0.21563
## Year2003           0.1087      0.2248      0.48  0.62876
## Year2004           0.1976      0.2096      0.94  0.34625
## Year2005           0.0230      0.2093      0.11  0.91269
## Year2006           0.0310      0.2085      0.15  0.88178
## Year2007           0.0526      0.2123      0.25  0.80437
## Year2008          -0.0361      0.2104     -0.17  0.86378
## Year2009          -0.0183      0.2096     -0.09  0.93054
## Year2010          -0.0754      0.2159     -0.35  0.72704
## Year2011          -0.0143      0.2241     -0.06  0.94926
## Year2012          -0.0892      0.2169     -0.41  0.68117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.244, Adjusted R-squared:  0.216
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 562 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0399 0.8630 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          1.63e-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 1.177 1 1.085
## LastAuthorFemale 1.167 1 1.081
## Year 1.243 16 1.007

```

Residuals from first and last author



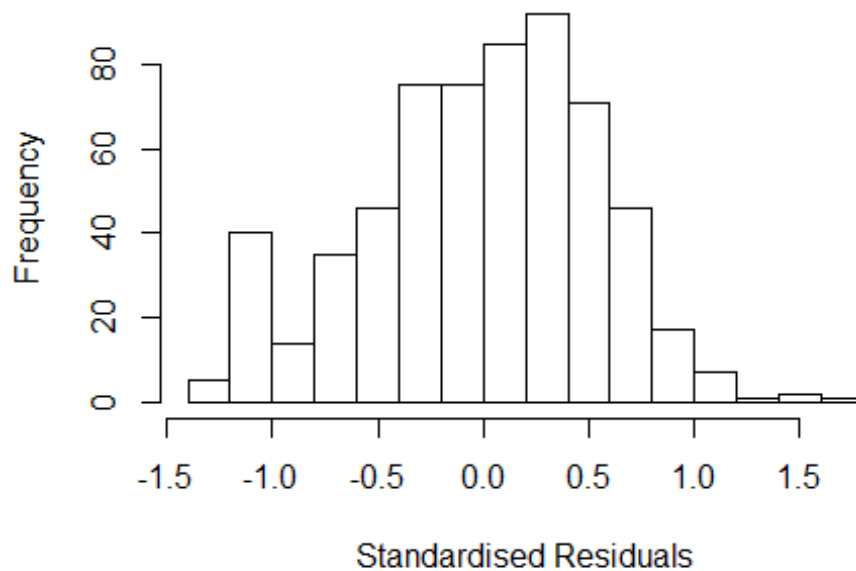
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3514 -0.3417 0.0255 0.3609 1.7869
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.350813 0.209716 6.44 2.5e-10 ***
## FirstAuthorFemale1 -0.073723 0.053721 -1.37 0.17
## LastAuthorFemale1 -0.215978 0.047666 -4.53 7.1e-06 ***
## Year1997 0.259131 0.255384 1.01 0.31
## Year1998 -0.040269 0.268154 -0.15 0.88
## Year1999 0.349562 0.267965 1.30 0.19
## Year2000 -0.161522 0.277625 -0.58 0.56
## Year2001 0.257830 0.239907 1.07 0.28
## Year2002 0.326618 0.228810 1.43 0.15
## Year2003 0.056483 0.227438 0.25 0.80
## Year2004 0.152103 0.222289 0.68 0.49
## Year2005 0.060742 0.219892 0.28 0.78
```

```

## Year2006          0.029901    0.216710    0.14    0.89
## Year2007          0.029521    0.221653    0.13    0.89
## Year2008         -0.092017    0.223114   -0.41    0.68
## Year2009         -0.093583    0.222399   -0.42    0.67
## Year2010         -0.097331    0.227492   -0.43    0.67
## Year2011          0.000564    0.238596    0.00    1.00
## Year2012         -0.073012    0.231173   -0.32    0.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.527
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0746
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 571 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.227  0.876  0.952  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      1.63e-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 1.155 1      1.075
## Year              1.155 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
## -1.3301 -0.3697 0.0437 0.3770 1.7086
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33015 0.20869 6.37 3.7e-10 ***
## FirstAuthorFemale1 -0.13134 0.05254 -2.50 0.013 *
## Year1997 0.22580 0.23972 0.94 0.347
## Year1998 -0.06516 0.27744 -0.23 0.814
## Year1999 0.39416 0.27332 1.44 0.150
## Year2000 -0.20861 0.28050 -0.74 0.457
## Year2001 0.27554 0.24921 1.11 0.269
## Year2002 0.27607 0.22858 1.21 0.228
## Year2003 0.02191 0.22869 0.10 0.924
## Year2004 0.11648 0.22341 0.52 0.602
## Year2005 0.00208 0.21985 0.01 0.992
## Year2006 -0.01861 0.21723 -0.09 0.932
```

```

## Year2007          -0.05889    0.22106   -0.27    0.790
## Year2008          -0.15145    0.22269   -0.68    0.497
## Year2009          -0.14542    0.22270   -0.65    0.514
## Year2010          -0.17470    0.22679   -0.77    0.441
## Year2011          -0.05237    0.23788   -0.22    0.826
## Year2012          -0.15505    0.23092   -0.67    0.502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.0694, Adjusted R-squared:  0.0428
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 562 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.308  0.874  0.950  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      1.63e-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 1.139 1      1.067
## Year      1.139 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
## -1.336 -0.344  0.032  0.364  1.775

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.33610    0.20667    6.46 2.1e-10 ***
## LastAuthorFemale1 -0.23271    0.04718   -4.93 1.1e-06 ***
## Year1997        0.25338    0.25590    0.99  0.32
## Year1998       -0.05724    0.26667   -0.21  0.83
## Year1999        0.33440    0.25887    1.29  0.20
## Year2000       -0.15604    0.27409   -0.57  0.57
## Year2001        0.21865    0.23519    0.93  0.35
## Year2002        0.29373    0.22695    1.29  0.20
## Year2003        0.02710    0.22484    0.12  0.90
## Year2004        0.12750    0.22009    0.58  0.56
## Year2005        0.02576    0.21600    0.12  0.91
## Year2006        0.00227    0.21311    0.01  0.99
## Year2007        0.00153    0.21874    0.01  0.99
## Year2008       -0.12239    0.21992   -0.56  0.58
## Year2009       -0.12501    0.21840   -0.57  0.57
## Year2010       -0.12904    0.22331   -0.58  0.56
## Year2011       -0.02458    0.23404   -0.11  0.92
## Year2012       -0.11077    0.22656   -0.49  0.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.098, Adjusted R-squared:  0.0722
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 561 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.874  0.951  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      1.63e-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: 612"
## [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]"
##
## 1997 1998 2002 2004 2005 2006 2009 2010 2011
##    3    1    1    3    3    1    2    2    1
##
## 1997 1998 2002 2004 2005 2006 2009 2010 2011
##    1    1    1    2    3    1    1    2    1
##
## 1997 1998 2002 2004 2005 2006 2009 2010 2011
##    1    1    1    2    3    1    1    2    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 6.48074069840786"
## [1] "Male first author team size 2018 geometric mean: 4"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 0.7
## 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: 4"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, 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"
## [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: 13"
## [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]"
##

```

```

## 2001 2002 2003 2004 2006 2007 2008 2009 2010 2011 2012
##      2      1      1      3      2      1      3      1      3      3      1
##
## 2001 2002 2003 2004 2006 2007 2008 2009 2010 2011 2012
##      1      1      1      2      2      1      3      1      3      3      1
##
## 2001 2002 2003 2004 2006 2007 2008 2009 2010 2011 2012
##      1      1      1      2      2      0      3      1      3      3      1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5.09145979004366"
## [1] "Male first author team size 2018 geometric mean: 3.46410161513775"

## 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 = 0.2
## 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: 3.7224194364084"

## 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 = 8, 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"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

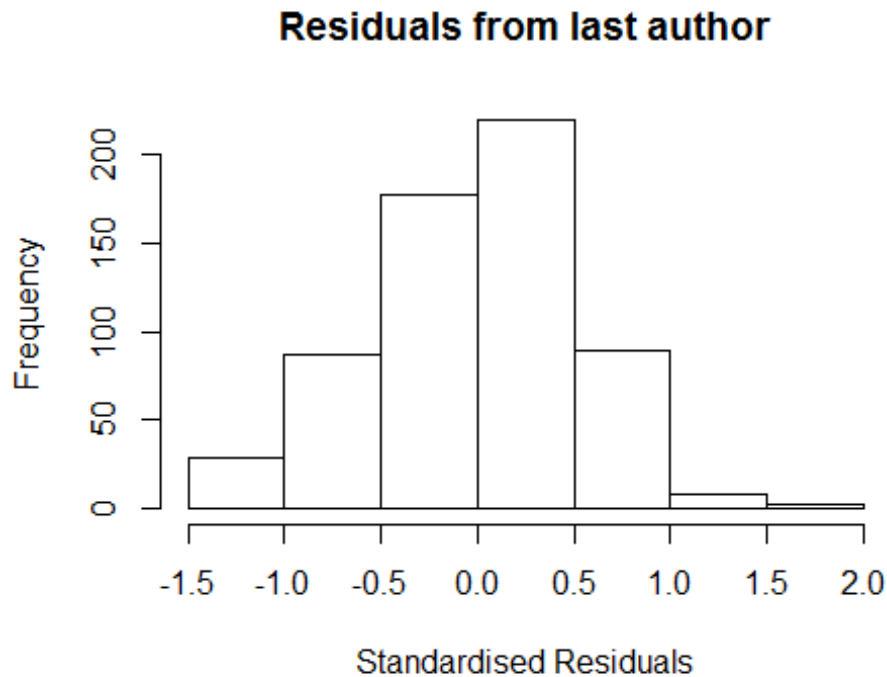
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

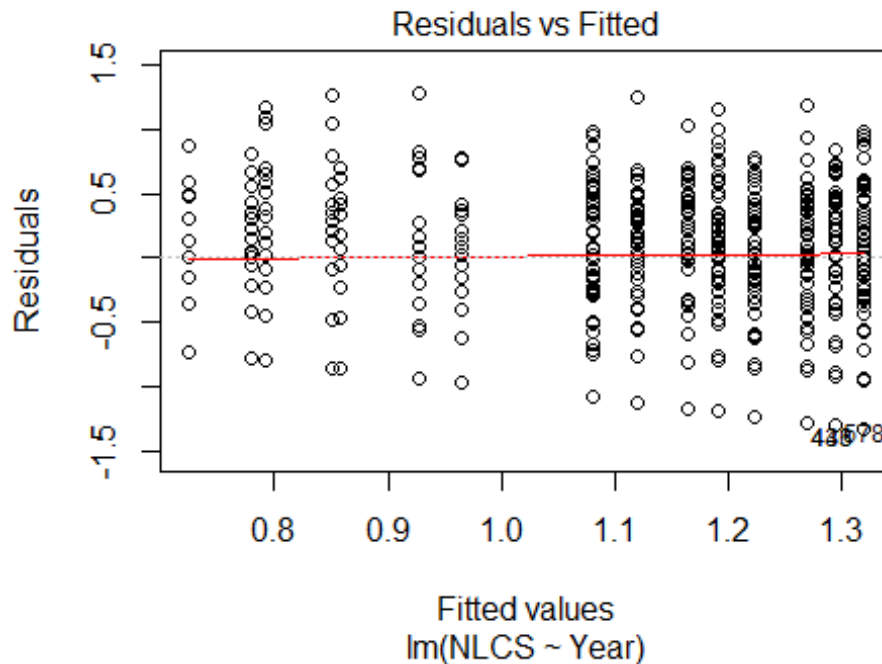
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

```



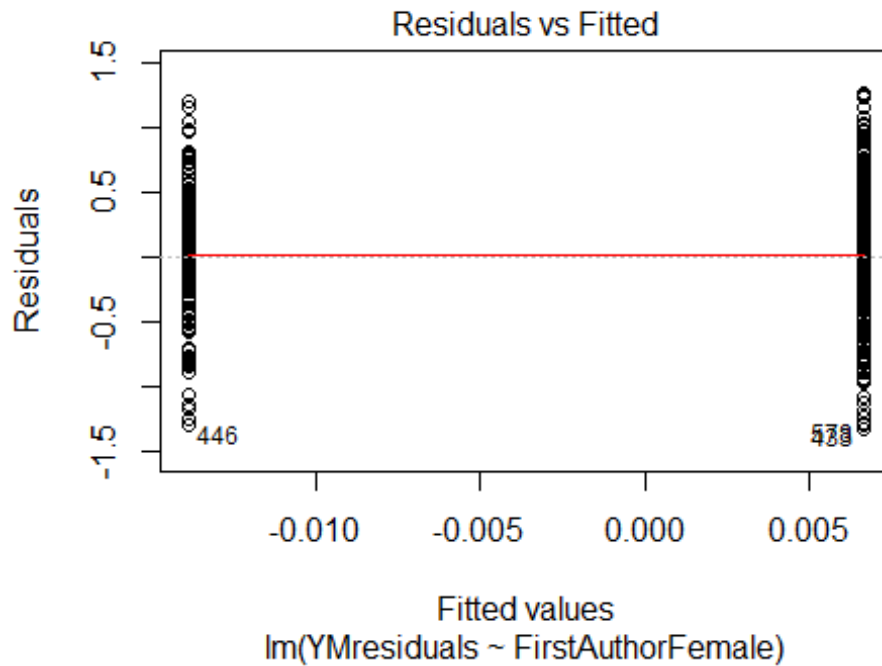
```
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 18"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2905"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   29   34   25   37   26   38   29   58   59   53   59   60   60   73   76
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   23   28   19   31   20   33   26   52   50   48   52   55   54   61   67
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   23   26   17   30   17   32   22   49   45   45   47   51   52   56   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 = 8.1, df = 14, p-value = 0.9
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0047, df = 1, p-value = 0.9

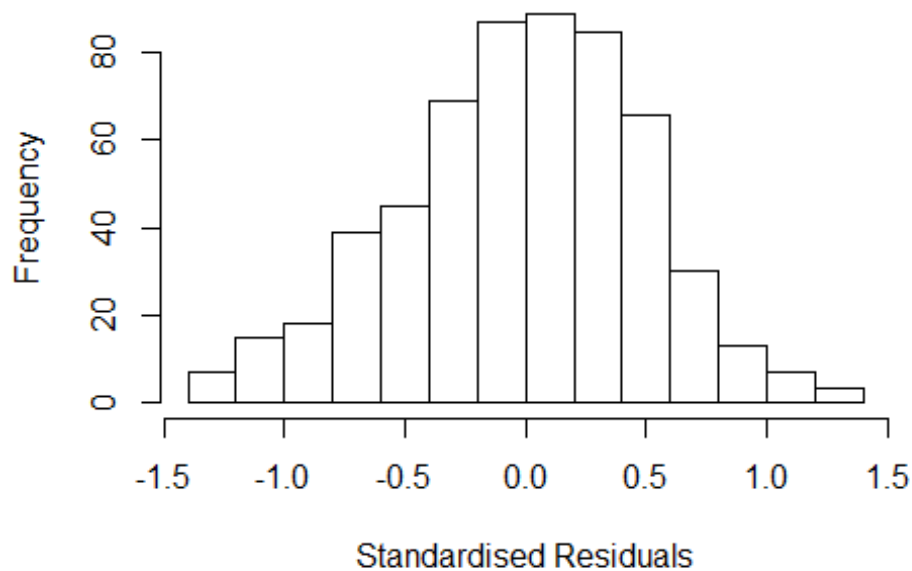
## [1] "Female first author team size 2018 geometric mean: 3.69978808157551"
## [1] "Male first author team size 2018 geometric mean: 4.00732331969539"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.70497645092122"
## [1] "Male last author team size 2018 geometric mean: 3.8871091118159"

## 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 = 710, 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 1.267 1          1.126
## LastAuthorFemale  1.135 1          1.065
## UniqueAuthors    2.077 4          1.096
## Year             2.256 14          1.029
```

Residuals from first and last author and team size



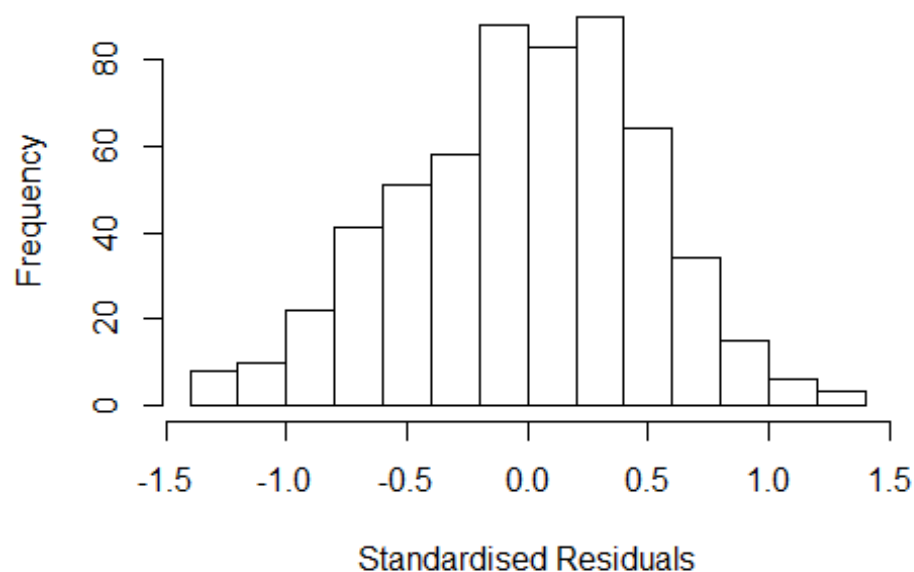
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3425 -0.3486 0.0111 0.3394 1.3084
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.73704 0.13732 5.37 1.2e-07 ***
## FirstAuthorFemale1 0.00882 0.05057 0.17 0.8616
## LastAuthorFemale1 0.07447 0.04575 1.63 0.1042
## UniqueAuthors2 0.02715 0.07842 0.35 0.7294
## UniqueAuthors3 0.14787 0.08234 1.80 0.0731 .
## UniqueAuthors4 0.14098 0.08929 1.58 0.1150
## UniqueAuthors5 0.18749 0.08645 2.17 0.0305 *
## Year1999 0.07034 0.15360 0.46 0.6472
## Year2000 -0.09141 0.20787 -0.44 0.6603
## Year2001 -0.04569 0.15288 -0.30 0.7652
```

```

## Year2002          0.14863    0.18100    0.82    0.4119
## Year2003         -0.13294    0.16396   -0.81    0.4179
## Year2004         -0.00209    0.18337   -0.01    0.9909
## Year2005          0.24377    0.15265    1.60    0.1109
## Year2006          0.30553    0.15737    1.94    0.0527 .
## Year2007          0.18561    0.16098    1.15    0.2494
## Year2008          0.42339    0.15693    2.70    0.0072 **
## Year2009          0.33205    0.14543    2.28    0.0228 *
## Year2010          0.44874    0.15475    2.90    0.0039 **
## Year2011          0.35803    0.15092    2.37    0.0180 *
## Year2012          0.28821    0.15229    1.89    0.0589 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.148, Adjusted R-squared:  0.117
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
##  57 weights are ~= 1. The remaining 516 ones are summarized as
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.453  0.867  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      1.75e-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 1.164 1      1.079
## LastAuthorFemale  1.090 1      1.044
## Year              1.176 14      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
## -1.3836 -0.3477 0.0193 0.3389 1.3105
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.78549 0.13283 5.91 5.9e-09 ***
## FirstAuthorFemale1 0.00574 0.04938 0.12 0.90747
## LastAuthorFemale1 0.06897 0.04574 1.51 0.13218
## Year1999 0.06589 0.15349 0.43 0.66788
## Year2000 -0.08698 0.19820 -0.44 0.66095
## Year2001 -0.04149 0.15158 -0.27 0.78441
## Year2002 0.19030 0.17048 1.12 0.26479
## Year2003 -0.09092 0.15997 -0.57 0.57001
## Year2004 0.09829 0.17685 0.56 0.57857
## Year2005 0.30799 0.14752 2.09 0.03727 *
## Year2006 0.34648 0.15576 2.22 0.02652 *
## Year2007 0.23663 0.15729 1.50 0.13304
```

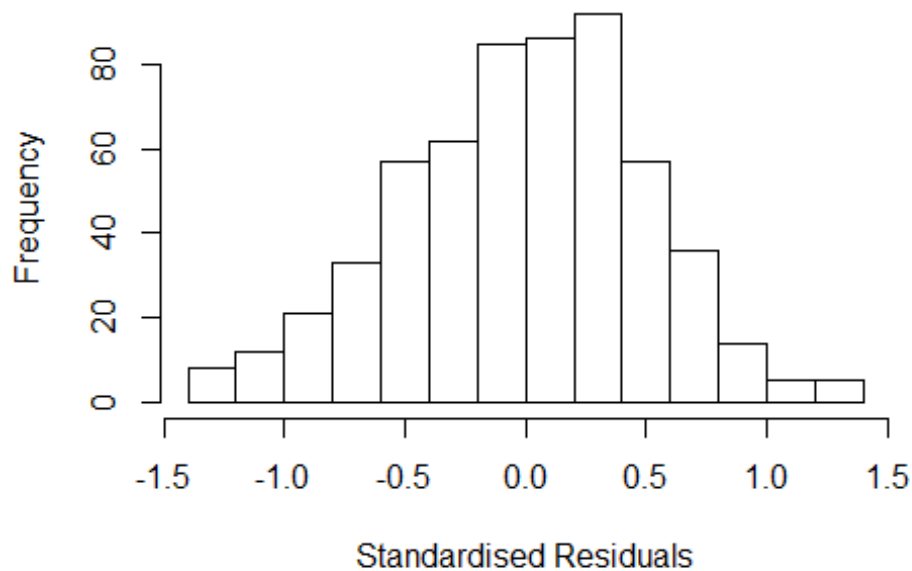


```

## Year2008          0.50651    0.15174    3.34  0.00090 ***
## Year2009          0.39025    0.14181    2.75  0.00612 **
## Year2010          0.52337    0.14919    3.51  0.00049 ***
## Year2011          0.43712    0.14256    3.07  0.00227 **
## Year2012          0.38153    0.14532    2.63  0.00889 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.135, Adjusted R-squared:  0.11
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 532 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.435  0.872  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      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 1.119 1          1.058
## Year              1.119 14          1.004

```

Residuals from first author



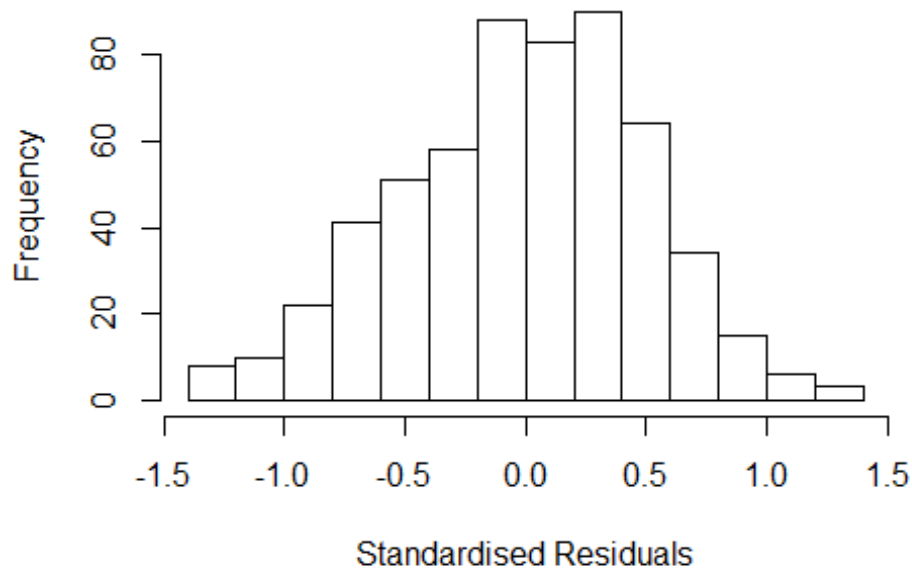
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3553 -0.3595 0.0253 0.3449 1.2793
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8100 0.1319 6.14 1.5e-09 ***
## FirstAuthorFemale1 0.0247 0.0486 0.51 0.61176
## Year1999 0.0612 0.1539 0.40 0.69093
## Year2000 -0.0877 0.2024 -0.43 0.66479
## Year2001 -0.0374 0.1514 -0.25 0.80497
## Year2002 0.1911 0.1736 1.10 0.27129
## Year2003 -0.0898 0.1633 -0.55 0.58249
## Year2004 0.0998 0.1787 0.56 0.57670
## Year2005 0.3113 0.1483 2.10 0.03628 *
## Year2006 0.3452 0.1571 2.20 0.02847 *
## Year2007 0.2460 0.1593 1.54 0.12314
## Year2008 0.5070 0.1539 3.29 0.00105 **
```

```

## Year2009          0.3964      0.1428      2.78  0.00569 **
## Year2010          0.5206      0.1494      3.48  0.00053 ***
## Year2011          0.4388      0.1436      3.06  0.00235 **
## Year2012          0.3884      0.1471      2.64  0.00851 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.131, Adjusted R-squared:  0.108
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 526 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.451  0.874  0.951  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      1.75e-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 1.051 1          1.025
## Year            1.051 14          1.002

```

Residuals from last author



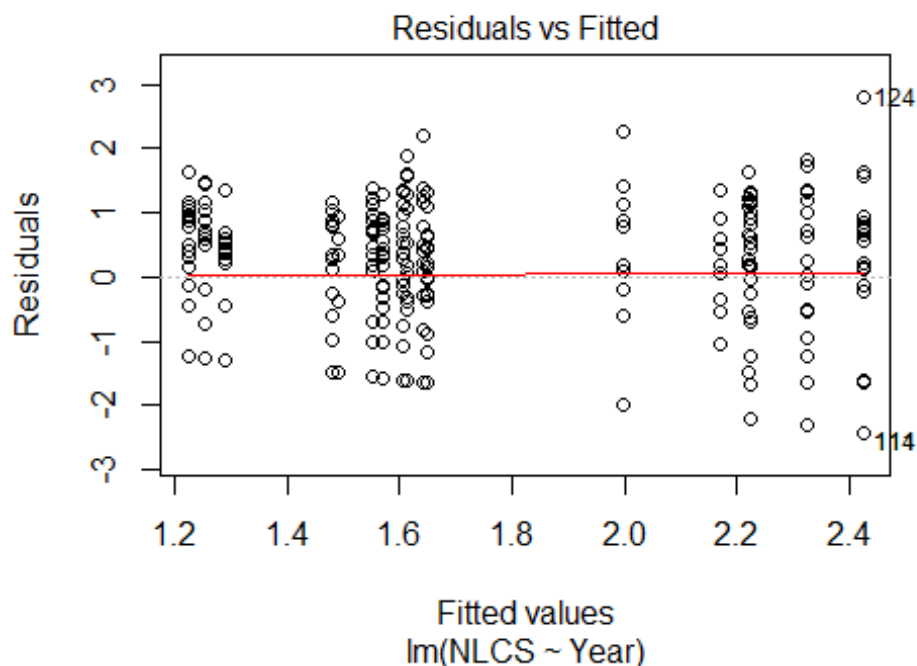
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3828 -0.3452 0.0203 0.3406 1.3138
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7873 0.1310 6.01 3.4e-09 ***
## LastAuthorFemale1 0.0703 0.0450 1.56 0.11894
## Year1999 0.0667 0.1535 0.43 0.66393
## Year2000 -0.0877 0.1978 -0.44 0.65772
## Year2001 -0.0406 0.1512 -0.27 0.78855
## Year2002 0.1914 0.1703 1.12 0.26137
## Year2003 -0.0901 0.1602 -0.56 0.57384
## Year2004 0.0989 0.1769 0.56 0.57646
## Year2005 0.3093 0.1473 2.10 0.03620 *
## Year2006 0.3481 0.1549 2.25 0.02503 *
## Year2007 0.2381 0.1567 1.52 0.12916
## Year2008 0.5081 0.1517 3.35 0.00086 ***
```

```

## Year2009          0.3922      0.1407      2.79  0.00549 **
## Year2010          0.5251      0.1488      3.53  0.00045 ***
## Year2011          0.4386      0.1418      3.09  0.00208 **
## Year2012          0.3830      0.1449      2.64  0.00843 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.135, Adjusted R-squared:  0.112
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 530 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.434  0.869  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      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: 573"
## [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
##    5   16   16   17   25   25   24   19   26   24   19   17   24   17   16
## 2011 2012
##   20   27
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5   10   13   11   20   22   21   18   23   21   17   16   23   16   15
## 2011 2012
##   17   21
##

```

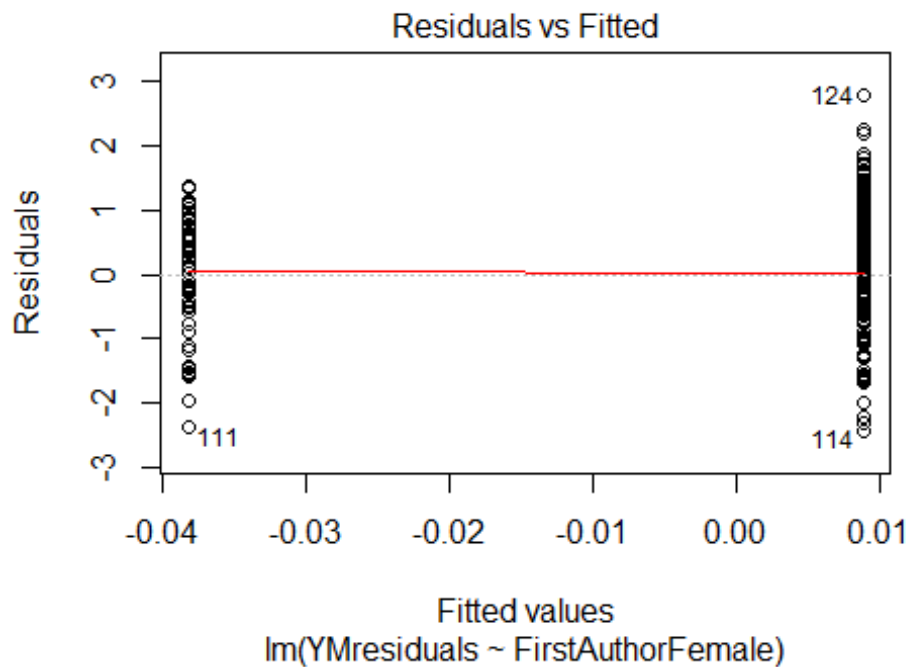
```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5   10   13   10   20   21   20   18   21   20   16   16   23   16   14
## 2011 2012
##   15   17
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity 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 = 1.9, df = 1, p-value = 0.2
##
## [1] "Female first author team size 2018 geometric mean: 2.83737610906523"
## [1] "Male first author team size 2018 geometric mean: 5.46037003054596"
##
## 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 = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.07700208712258"
## [1] "Male last author team size 2018 geometric mean: 4.50397764970732"

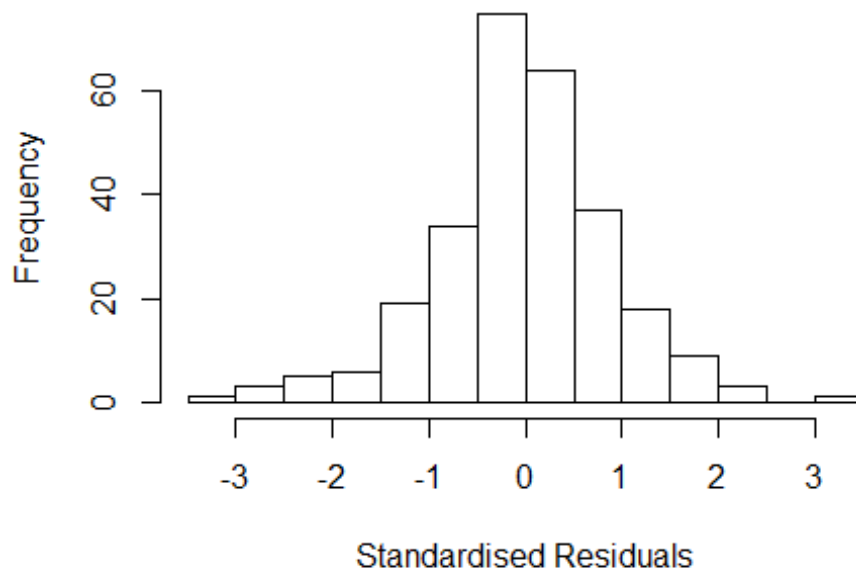
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 50, 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	1.666	1	1.291
LastAuthorFemale	1.962	1	1.401
UniqueAuthors	6.530	4	1.264
Year	10.037	16	1.075

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
## 10   0031278247 0.715 1997    2906     1   -2.597
## 35   0032130460 4.254 1998    2906     1    3.262
## 67   33749058071 0.000 2000    2906     2   -2.967
## 111  0036688583 0.000 2002    2906     2   -2.754
## 164  3543107806 0.000 2004    2906     2   -3.056
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0557 -0.4912 -0.0571  0.4859  3.2618
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8369    0.4977   1.68   0.094 .
## FirstAuthorFemale1  0.1785    0.1740   1.03   0.306
## LastAuthorFemale1 -0.0771    0.1423  -0.54   0.589
## UniqueAuthors2    1.5569    0.1943   8.01 4.2e-14 ***
## UniqueAuthors3    1.3382    0.2088   6.41 7.2e-10 ***
## UniqueAuthors4    1.6348    0.2373   6.89 4.4e-11 ***
```

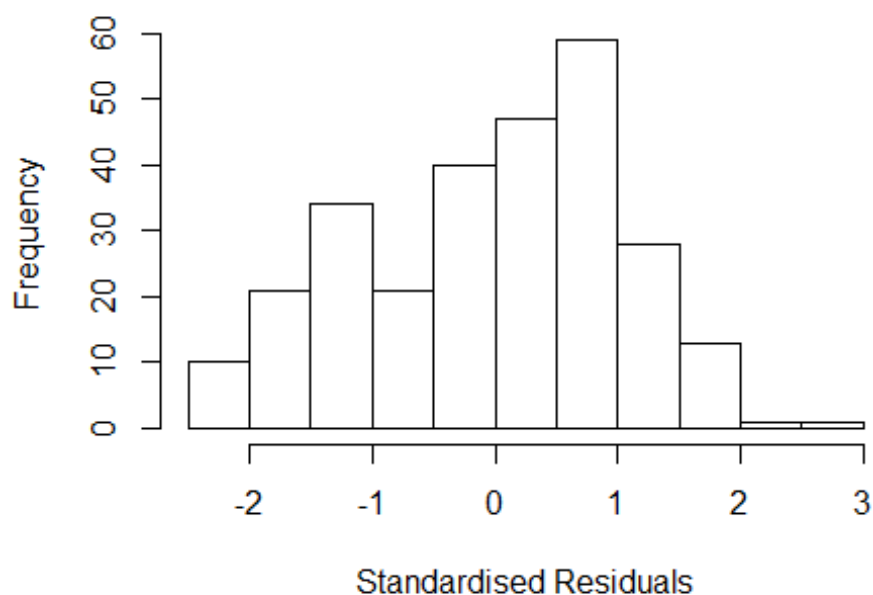


```

## UniqueAuthors5      1.3753      0.2049      6.71 1.3e-10 ***
## Year1997             1.0352      0.7852      1.32  0.189
## Year1998             0.0538      0.5114      0.11  0.916
## Year1999             0.2348      0.5938      0.40  0.693
## Year2000             0.6909      0.5013      1.38  0.169
## Year2001            -0.4092      0.4979     -0.82  0.412
## Year2002             0.4370      0.4789      0.91  0.362
## Year2003            -0.3653      0.4794     -0.76  0.447
## Year2004             0.5604      0.4653      1.20  0.230
## Year2005            -0.4976      0.4097     -1.21  0.226
## Year2006            -0.6132      0.4651     -1.32  0.189
## Year2007            -0.3844      0.4606     -0.83  0.405
## Year2008            -0.5577      0.4114     -1.36  0.176
## Year2009            -0.7991      0.4271     -1.87  0.062 .
## Year2010            -0.5323      0.4931     -1.08  0.281
## Year2011            -0.5816      0.4365     -1.33  0.184
## Year2012            -0.4471      0.4516     -0.99  0.323
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.728
## Multiple R-squared:  0.513, Adjusted R-squared:  0.471
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0072 0.8370 0.9520 0.8700 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      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 1.341 1 1.158
## LastAuthorFemale 1.422 1 1.192
## Year 1.528 16 1.013

```

Residuals from first and last author



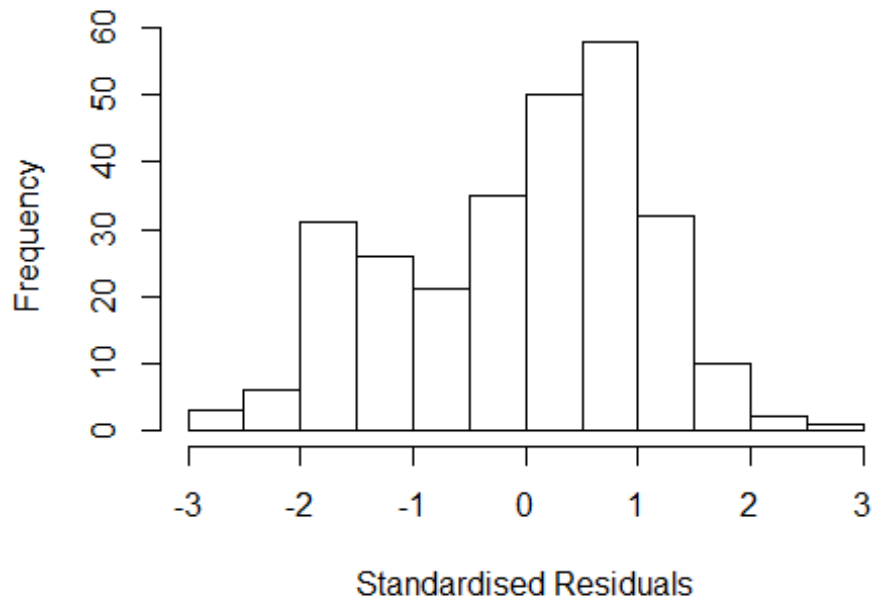
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 124 0036011007 5.222 2002      2906      1      2.816
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.426 -0.898  0.129  0.773  2.816
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.62577    0.44807   3.63  0.00034 ***
## FirstAuthorFemale1  0.26246    0.17881   1.47  0.14338
## LastAuthorFemale1 -0.48863    0.14613  -3.34  0.00095 ***
## Year1997         0.76064    0.57956   1.31  0.19054
## Year1998         0.62426    0.57438   1.09  0.27814
## Year1999         0.61413    0.43889   1.40  0.16294
## Year2000         0.77709    0.50360   1.54  0.12405
## Year2001         0.12297    0.47234   0.26  0.79481
## Year2002         1.00685    0.53498   1.88  0.06096 .
## Year2003         0.18289    0.49365   0.37  0.71133
## Year2004         1.02605    0.44945   2.28  0.02326 *
## Year2005        -0.00305    0.45785  -0.01  0.99470
```

```

## Year2006      -0.25740      0.49190      -0.52      0.60123
## Year2007      0.11627      0.50397      0.23      0.81772
## Year2008     -0.27910      0.44165     -0.63      0.52797
## Year2009     -0.24078      0.44781     -0.54      0.59127
## Year2010     -0.04216      0.49118     -0.09      0.93167
## Year2011      0.20210      0.44188      0.46      0.64780
## Year2012     -0.01048      0.46216     -0.02      0.98193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 1.12
## Multiple R-squared:  0.155, Adjusted R-squared:  0.0959
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 256 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.506  0.865   0.946   0.918   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      3.64e-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 1.225 1      1.107
## Year              1.225 16      1.006

```

Residuals from first author



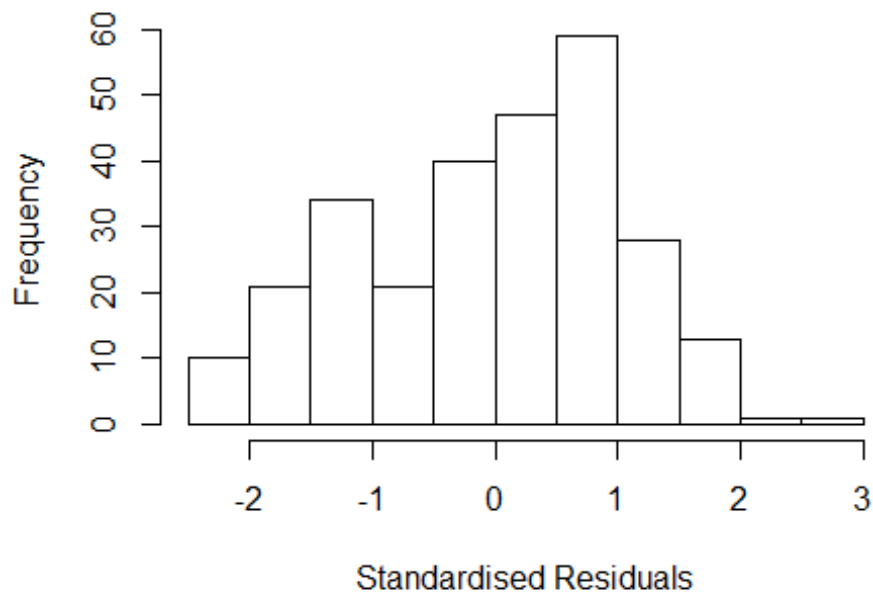
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 124 0036011007 5.222 2002      2906      1      2.816
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.502 -0.949  0.187  0.763  2.760
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.47470    0.44050   3.35  0.00094 ***
## FirstAuthorFemale1 0.06185    0.16958   0.36  0.71564
## Year1997          0.72954    0.58611   1.24  0.21437
## Year1998          0.52690    0.56810   0.93  0.35455
## Year1999          0.50213    0.44911   1.12  0.26459
## Year2000          0.82105    0.49443   1.66  0.09801 .
## Year2001          0.06050    0.47057   0.13  0.89781
## Year2002          0.92554    0.53426   1.73  0.08441 .
## Year2003          0.08055    0.49518   0.16  0.87091
## Year2004          0.96528    0.45780   2.11  0.03595 *
## Year2005          0.00968    0.46867   0.02  0.98354
## Year2006         -0.32012    0.50750  -0.63  0.52874
```

```

## Year2007          0.12717      0.52102      0.24  0.80736
## Year2008          -0.29450      0.45789     -0.64  0.52070
## Year2009          -0.21661      0.46173     -0.47  0.63938
## Year2010           0.08628      0.48342      0.18  0.85848
## Year2011           0.13493      0.45306      0.30  0.76608
## Year2012           0.04565      0.46427      0.10  0.92174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 1.11
## Multiple R-squared:  0.128, Adjusted R-squared:  0.0708
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 262 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.516  0.875   0.944   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      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 1.274 1          1.129
## Year              1.274 16          1.008

```

Residuals from last author



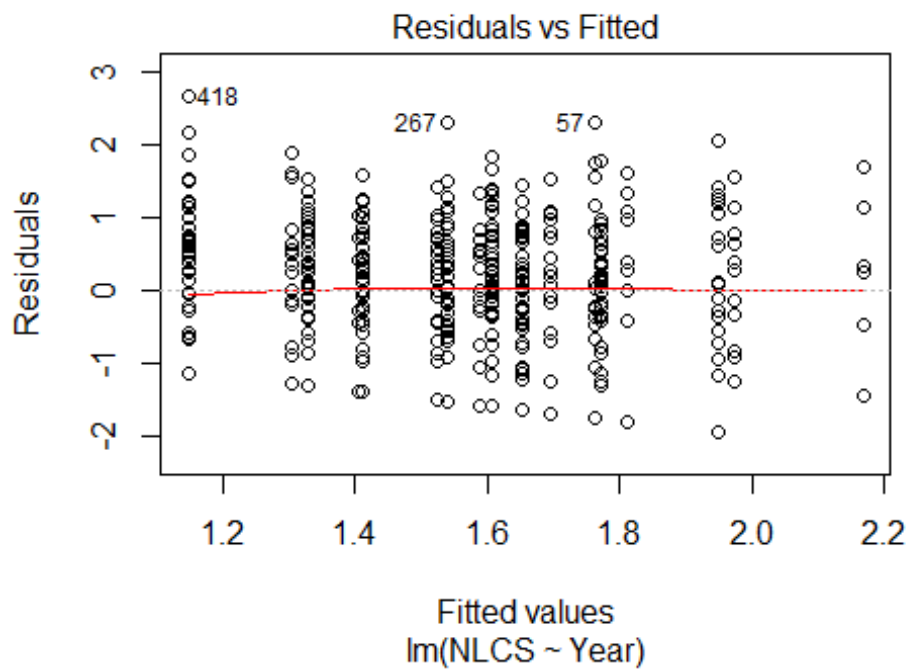
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 124 0036011007 5.222 2002      2906      1      2.816
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.425 -0.889  0.101  0.785  2.844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.8077     0.4483   4.03 7.3e-05 ***
## LastAuthorFemale1 -0.4084     0.1396  -2.93  0.0037 **
## Year1997          0.6941     0.5944   1.17  0.2440
## Year1998          0.6031     0.6031   1.00  0.3183
## Year1999          0.5971     0.4637   1.29  0.1989
## Year2000          0.7396     0.5325   1.39  0.1660
## Year2001          0.0743     0.4876   0.15  0.8790
## Year2002          0.9783     0.5583   1.75  0.0809 .
## Year2003          0.1473     0.5198   0.28  0.7770
## Year2004          1.0252     0.4778   2.15  0.0328 *
## Year2005          0.0160     0.4903   0.03  0.9740
## Year2006         -0.2784     0.5108  -0.55  0.5862
```

```

## Year2007          0.0885      0.5351      0.17      0.8687
## Year2008          -0.2814      0.4717     -0.60      0.5514
## Year2009          -0.2601      0.4757     -0.55      0.5850
## Year2010          -0.1031      0.5201     -0.20      0.8430
## Year2011           0.1382      0.4656      0.30      0.7668
## Year2012          -0.0253      0.4894     -0.05      0.9588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 1.12
## Multiple R-squared:  0.15,   Adjusted R-squared:  0.0938
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.498  0.868  0.944  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      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 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
##    6   12   15   17   22   30   27   22   23   30   40   37   64   49   67
## 2011 2012
##   73   62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    7   12   13   17   24   21   22   20   27   34   35   62   44   61
## 2011 2012

```

```
## 65 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 7 12 12 17 24 19 21 18 25 33 33 52 40 57
## 2011 2012
## 63 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 = 18, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1

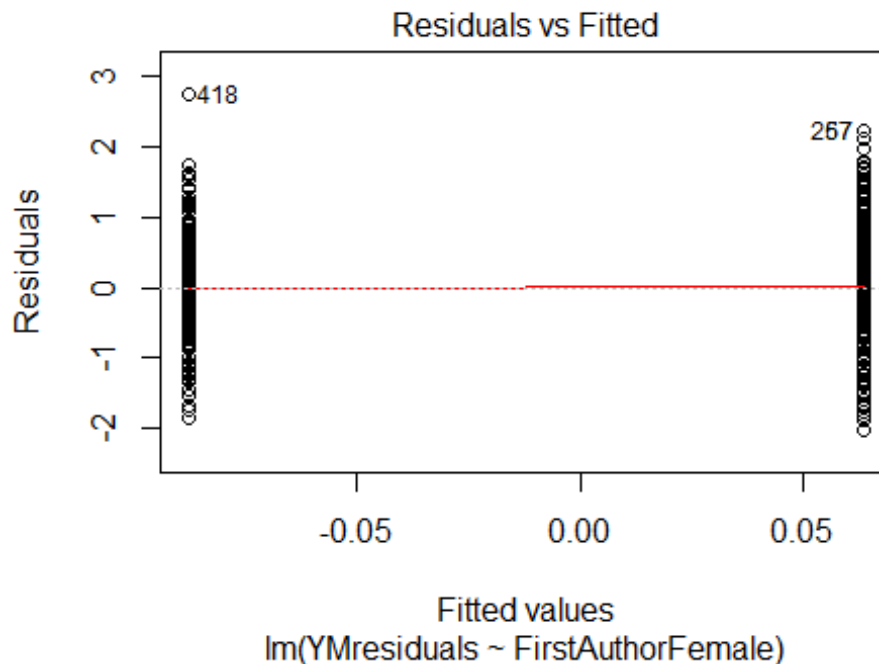
## [1] "Female first author team size 2018 geometric mean: 3.58204802702754"
## [1] "Male first author team size 2018 geometric mean: 4.65596466758839"

## 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 = 290, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.43473536510974"
## [1] "Male last author team size 2018 geometric mean: 4.70908860526264"

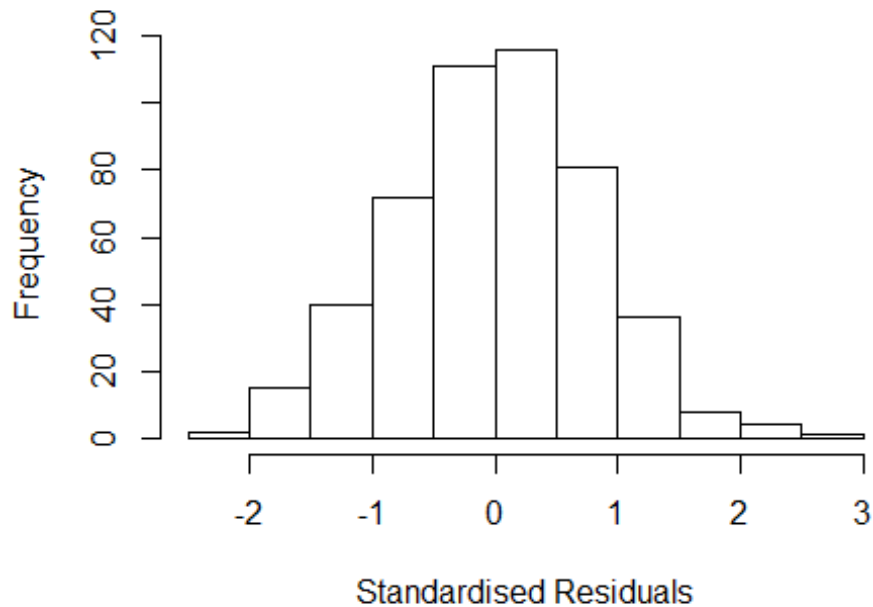
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.02
## 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	1.604	1	1.267
LastAuthorFemale	1.328	1	1.152
UniqueAuthors	1.605	4	1.061
Year	2.440	16	1.028

Residuals from first and last author and team size



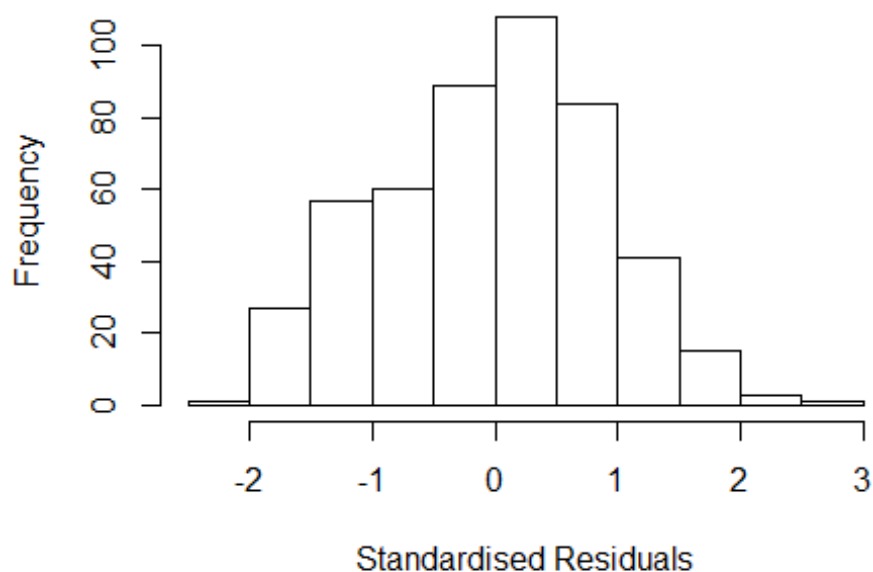
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 78649579593 3.803 2010      2711      2      2.675
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1842 -0.5391  0.0207  0.5436  2.6754
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0647    0.2941   3.62 0.00033 ***
## FirstAuthorFemale1  0.1810    0.0902   2.01 0.04532 *
## LastAuthorFemale1 -0.0153    0.0832  -0.18 0.85416
## UniqueAuthors2     0.7451    0.1207   6.17 1.5e-09 ***
## UniqueAuthors3     0.7806    0.1182   6.61 1.1e-10 ***
## UniqueAuthors4     1.0000    0.1446   6.92 1.5e-11 ***
## UniqueAuthors5     1.0313    0.1117   9.23 < 2e-16 ***
## Year1997          0.8305    0.6646   1.25 0.21210
## Year1998          0.1572    0.4402   0.36 0.72117
## Year1999          0.1120    0.3976   0.28 0.77824
```

```

## Year2000      0.1078      0.3979      0.27  0.78648
## Year2001     -0.2823      0.3641     -0.78  0.43858
## Year2002      0.0615      0.4424      0.14  0.88942
## Year2003     -0.4699      0.3504     -1.34  0.18056
## Year2004      0.2087      0.3730      0.56  0.57616
## Year2005     -0.2767      0.3180     -0.87  0.38459
## Year2006     -0.0104      0.3191     -0.03  0.97408
## Year2007     -0.2066      0.3260     -0.63  0.52644
## Year2008     -0.3721      0.3113     -1.20  0.23265
## Year2009     -0.3669      0.3179     -1.15  0.24907
## Year2010     -0.7177      0.3142     -2.28  0.02284 *
## Year2011     -0.2524      0.3141     -0.80  0.42201
## Year2012     -0.3401      0.3201     -1.06  0.28860
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.795
## Multiple R-squared:  0.252, Adjusted R-squared:  0.216
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 448 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.234  0.867  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      2.06e-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 1.454 1      1.206
## LastAuthorFemale  1.244 1      1.115
## Year              1.445 16      1.012

```

Residuals from first and last author



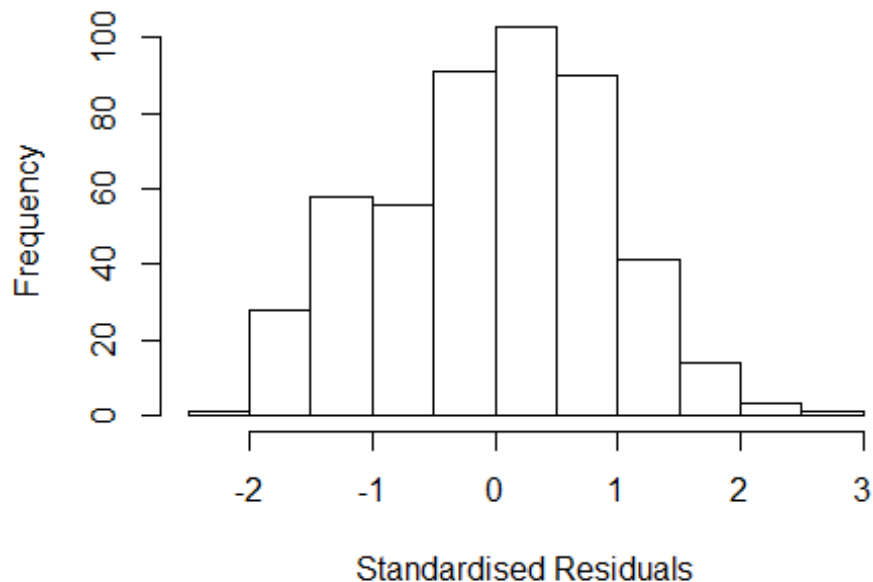
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 78649579593 3.803 2010      2711      2      2.771
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0042 -0.6538  0.0584  0.6620  2.7711
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.33602    0.34215   3.90  0.00011 ***
## FirstAuthorFemale1  0.23443    0.09851   2.38  0.01772 *
## LastAuthorFemale1 -0.08308    0.09167  -0.91  0.36522
## Year1997          0.76288    0.62435   1.22  0.22237
## Year1998          0.45206    0.51799   0.87  0.38327
## Year1999          0.36416    0.40897   0.89  0.37370
## Year2000          0.31145    0.44892   0.69  0.48817
## Year2001          0.18148    0.38948   0.47  0.64147
## Year2002          0.51683    0.44562   1.16  0.24673
## Year2003         -0.15679    0.41025  -0.38  0.70250
## Year2004          0.33856    0.42737   0.79  0.42866
## Year2005          0.00438    0.37586   0.01  0.99071
```

```

## Year2006          0.37289      0.35579      1.05  0.29515
## Year2007          0.12935      0.37454      0.35  0.72999
## Year2008         -0.06742      0.35679     -0.19  0.85022
## Year2009         -0.02213      0.36161     -0.06  0.95123
## Year2010         -0.30410      0.36135     -0.84  0.40046
## Year2011          0.19159      0.35485      0.54  0.58951
## Year2012          0.21185      0.35304      0.60  0.54875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.954
## Multiple R-squared:  0.0763, Adjusted R-squared:  0.0407
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 442 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.379  0.869  0.949  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      2.06e-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 1.324 1      1.151
## Year              1.324 16      1.009

```

Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 78649579593 3.803 2010      2711      2      2.771
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.022 -0.637  0.055  0.652  2.787
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.32e+00   3.46e-01   3.83  0.00015 ***
## FirstAuthorFemale1 1.97e-01   9.46e-02   2.08  0.03782 *
## Year1997        7.54e-01   6.32e-01   1.19  0.23341
## Year1998        4.26e-01   5.20e-01   0.82  0.41332
## Year1999        3.41e-01   4.10e-01   0.83  0.40618
## Year2000        3.10e-01   4.51e-01   0.69  0.49200
## Year2001        1.66e-01   3.93e-01   0.42  0.67288
## Year2002        5.01e-01   4.48e-01   1.12  0.26383
## Year2003       -1.83e-01   4.16e-01  -0.44  0.66001
## Year2004        3.25e-01   4.34e-01   0.75  0.45400
## Year2005       -7.54e-05   3.81e-01   0.00  0.99984
## Year2006        3.54e-01   3.61e-01   0.98  0.32696
```

```

## Year2007          1.19e-01   3.79e-01    0.31  0.75388
## Year2008          -7.88e-02   3.62e-01   -0.22  0.82754
## Year2009          -3.28e-02   3.67e-01   -0.09  0.92887
## Year2010          -3.08e-01   3.66e-01   -0.84  0.40011
## Year2011          1.93e-01   3.60e-01    0.54  0.59138
## Year2012          2.08e-01   3.58e-01    0.58  0.56178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.951
## Multiple R-squared:  0.0753, Adjusted R-squared:  0.0417
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.371  0.873  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      2.06e-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 1.147 1      1.071
## Year      1.147 16      1.004

## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 418 78649579593 3.803 2010      2711      2      2.771
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.931 -0.637  0.022  0.660  2.732

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4239    0.3624   3.93 9.8e-05 ***
## LastAuthorFemale1  0.0173    0.0891   0.19  0.85
## Year1997          0.7352    0.6262   1.17  0.24
## Year1998          0.4790    0.5374   0.89  0.37
## Year1999          0.3985    0.4249   0.94  0.35
## Year2000          0.2562    0.4673   0.55  0.58
## Year2001          0.1913    0.4029   0.47  0.64
## Year2002          0.4902    0.4640   1.06  0.29
## Year2003         -0.1905    0.4261  -0.45  0.65
## Year2004          0.3470    0.4448   0.78  0.44
## Year2005          0.0456    0.3985   0.11  0.91
## Year2006          0.3293    0.3786   0.87  0.38
## Year2007          0.1364    0.3977   0.34  0.73
## Year2008         -0.0770    0.3793  -0.20  0.84
## Year2009         -0.0240    0.3831  -0.06  0.95
## Year2010         -0.3531    0.3832  -0.92  0.36
## Year2011          0.1934    0.3774   0.51  0.61
## Year2012          0.2398    0.3753   0.64  0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.953
## Multiple R-squared:  0.0652, Adjusted R-squared:  0.0312
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 442 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.391  0.873   0.948   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      2.06e-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: 486"
## [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]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    2    3    2    4    7    1    2    2    3    4    7    6    2    9    8
## 2012
##    11
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    0    1    0    4    0    0    1    0    0    5    5    1    5    5
## 2012
##     5
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    0    1    0    3    0    0    1    0    0    4    3    1    5    5
## 2012
##     5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.86325050653529"
## [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

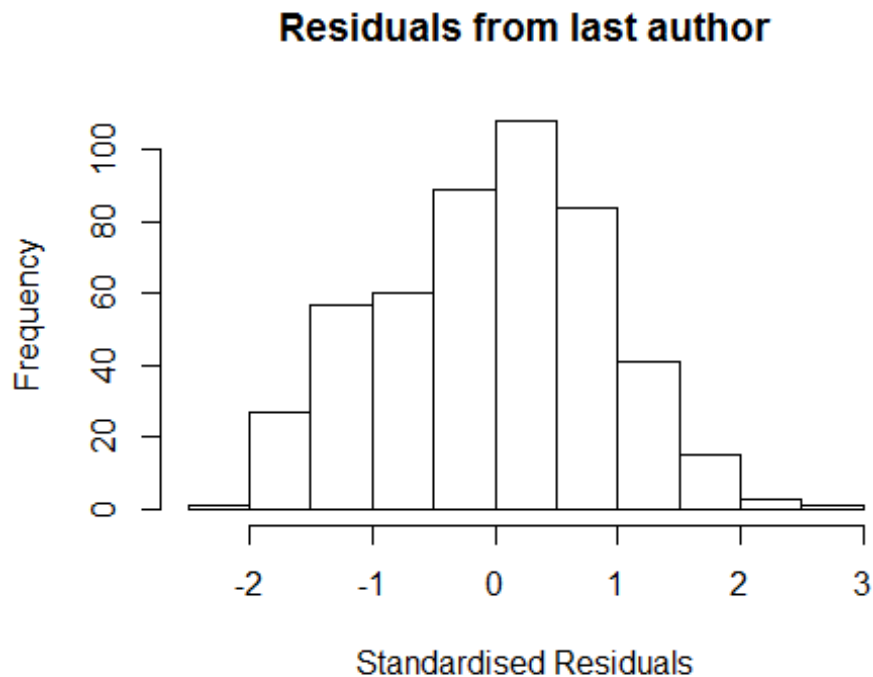
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.64939976115097"
## [1] "Male last author team size 2018 geometric mean: 3.66284150148471"

## 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 = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

```

```
## 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"
##                               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -4.242e+14  1             NaN
## LastAuthorFemale  1.810e+15  1          42544107
## Year              -1.247e+30  8             NaN

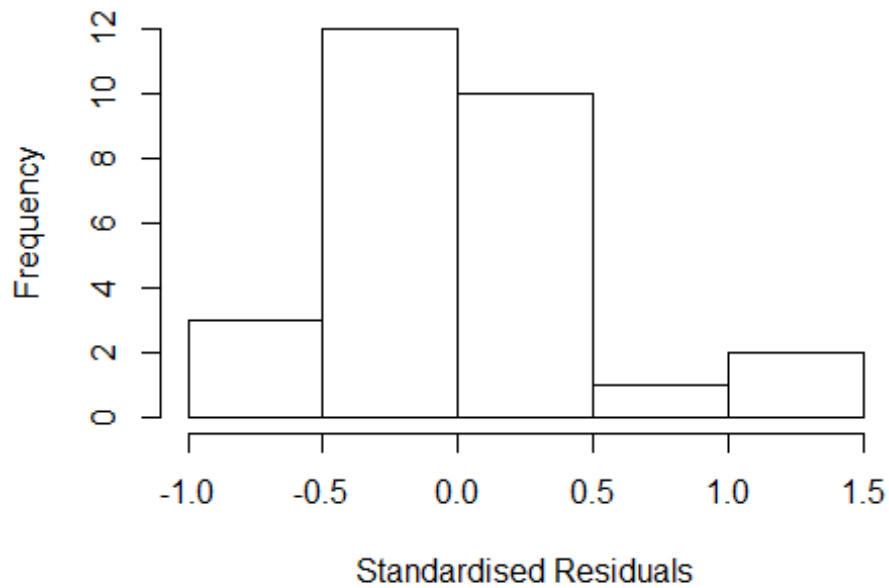
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.79e-01 -2.20e-01  9.44e-16  1.81e-01  1.08e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.728      0.301    2.42  0.02708 *
## FirstAuthorFemale1  0.204      0.151    1.35  0.19511
```

```

## LastAuthorFemale1    -0.932      0.186    -5.02  0.00011 ***
## Year2001             0.525      0.582     0.90  0.37921
## Year2004             0.597      0.301     1.98  0.06363 .
## Year2007             0.674      0.315     2.14  0.04726 *
## Year2008             0.628      0.182     3.44  0.00310 **
## Year2009             0.514      0.186     2.77  0.01320 *
## Year2010             0.853      0.142     6.02  1.4e-05 ***
## Year2011             0.873      0.134     6.51  5.4e-06 ***
## Year2012             0.220      0.255     0.86  0.40093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.623, Adjusted R-squared:  0.402
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 24 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.467  0.913  0.974   0.908   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      3.57e-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 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

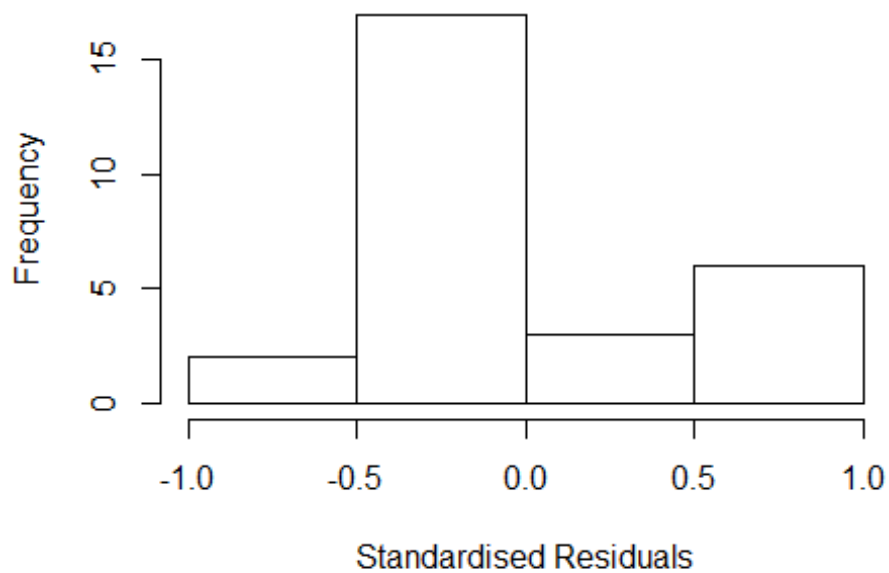
```

Residuals from first and last author



##		GVIF	Df	$GVIF^{(1/(2*Df))}$
##	FirstAuthorFemale	NaN	1	NaN
##	Year	NaN	8	NaN

Residuals from first author

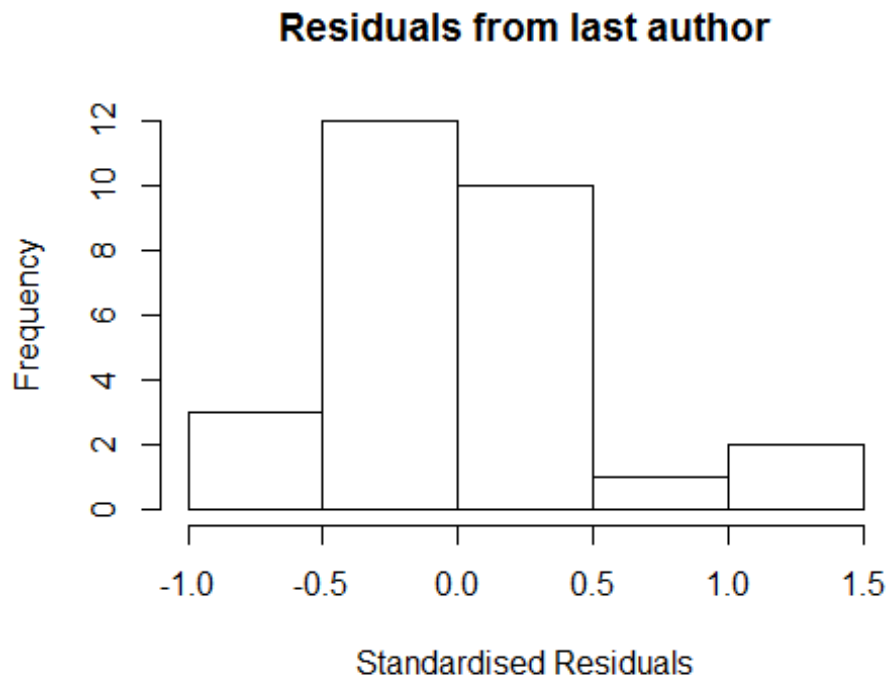


```

## [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
## -0.8628 -0.3691 -0.0408 0.3779 0.9812
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.048 0.439 -0.11 0.9141
## FirstAuthorFemale1 0.048 0.439 0.11 0.9141
## Year2001 0.610 0.482 1.27 0.2218
## Year2004 1.373 0.439 3.13 0.0058 **
## Year2007 0.863 0.475 1.81 0.0862 .
## Year2008 0.837 0.349 2.39 0.0277 *
## Year2009 1.446 0.000 Inf <2e-16 ***
## Year2010 1.058 0.396 2.67 0.0156 *
## Year2011 0.983 0.237 4.15 0.0006 ***
## Year2012 0.476 0.401 1.19 0.2502
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared: 0.231, Adjusted R-squared: -0.154
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 22 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.738 0.816 0.936 0.903 0.968 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 3.57e-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 1.126e+16 1 1.061e+08
## Year            1.126e+16 8 1.007e+01
```



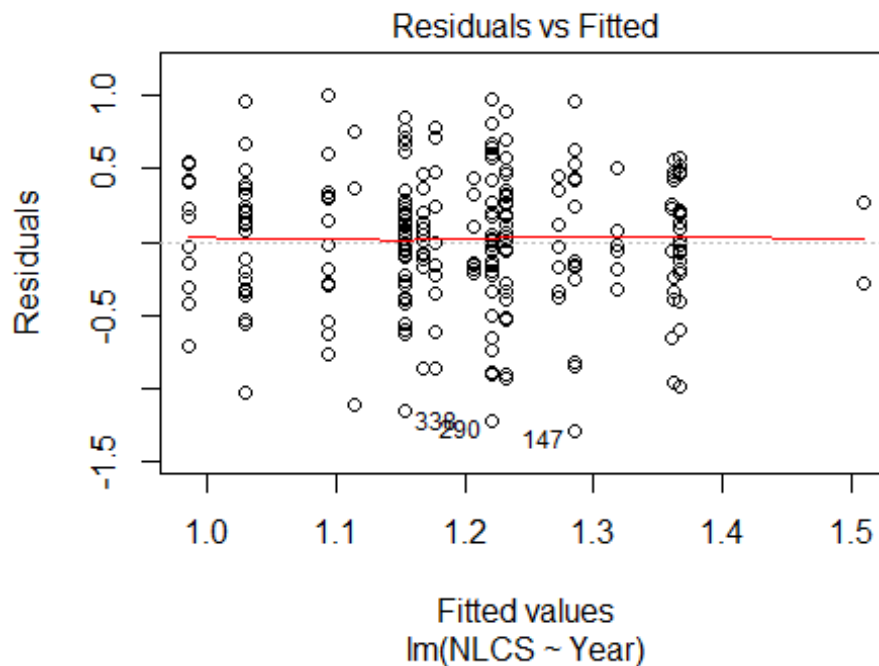
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.74e-01 -1.96e-01 -7.22e-16  1.87e-01  1.11e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.928      0.233    3.99  0.00087 ***
## LastAuthorFemale1 -0.928      0.233   -3.99  0.00087 ***
## Year2001          0.470      0.609    0.77  0.45048
## Year2004          0.397      0.233    1.71  0.10525
## Year2007          0.601      0.320    1.88  0.07672 .
```

```

## Year2008          0.498      0.132      3.77  0.00141 **
## Year2009          0.518      0.233      2.23  0.03907 *
## Year2010          0.748      0.157      4.77  0.00015 ***
## Year2011          0.833      0.123      6.78  2.4e-06 ***
## Year2012          0.196      0.254      0.77  0.45191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.644, Adjusted R-squared:  0.467
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 21 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.368  0.864  0.959  0.873  0.976  0.990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+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-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: 28"
## [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
##    4    9    9   10   12   24   14   12   10   10   18   17   23   33   34
## 2011 2012
##   44   51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    7    5    2    8   12   12   10    8    6   13   14   21   27   30
## 2011 2012
##   32   41

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    7    2    2    6   12   12    8    8    6    9   14   17   25   26
## 2011 2012
##   28   36
## [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
```

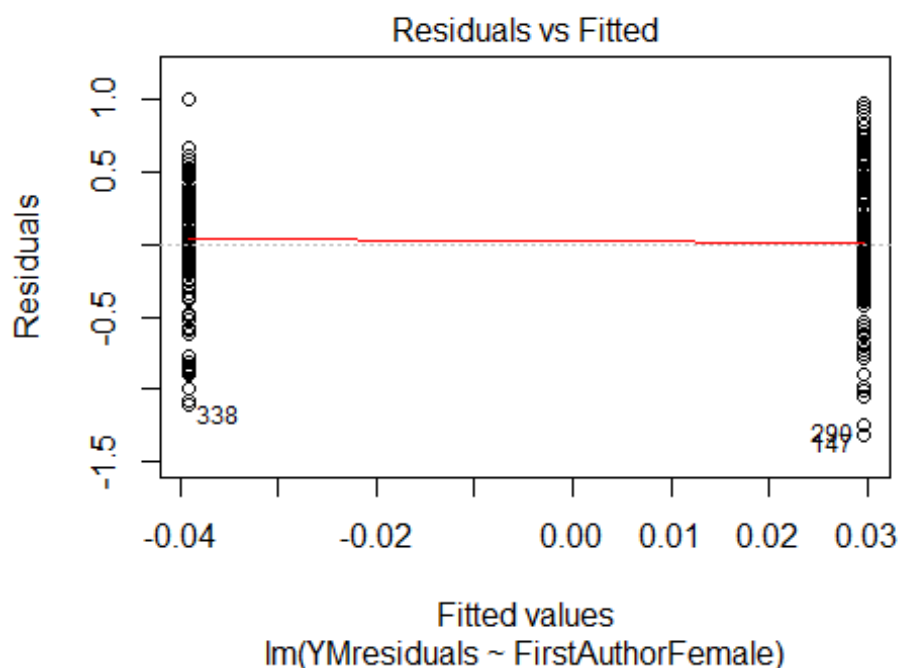


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.36, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 3.47075943428067"
## [1] "Male first author team size 2018 geometric mean: 4.79035262372259"
## 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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.40357370085871"
## [1] "Male last author team size 2018 geometric mean: 4.50966827552911"

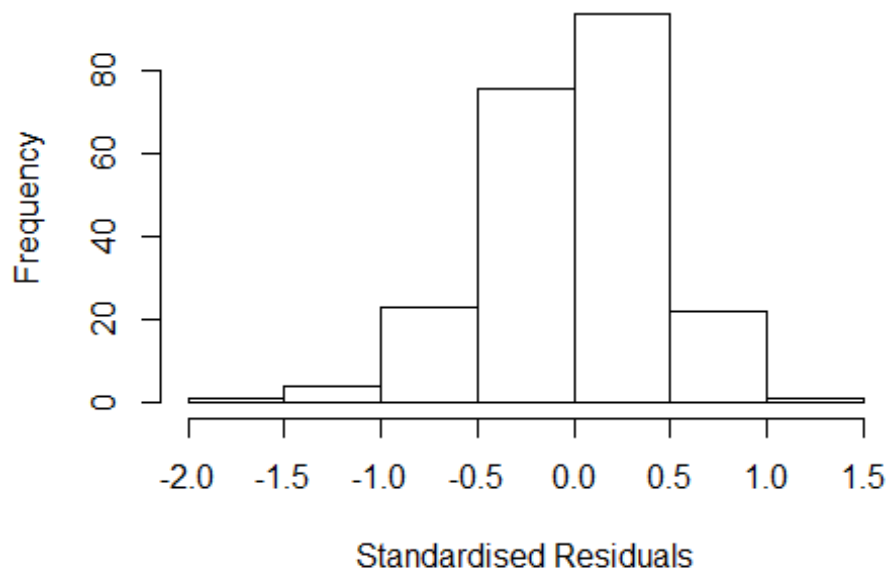
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, 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"
##
```

	GVI	F	Df	GVI ^{1/(2*Df)}
FirstAuthorFemale	2.921	1		1.709
LastAuthorFemale	2.046	1		1.430
UniqueAuthors	30.097	4		1.530
Year	68.470	16		1.141

Residuals from first and last author and team size



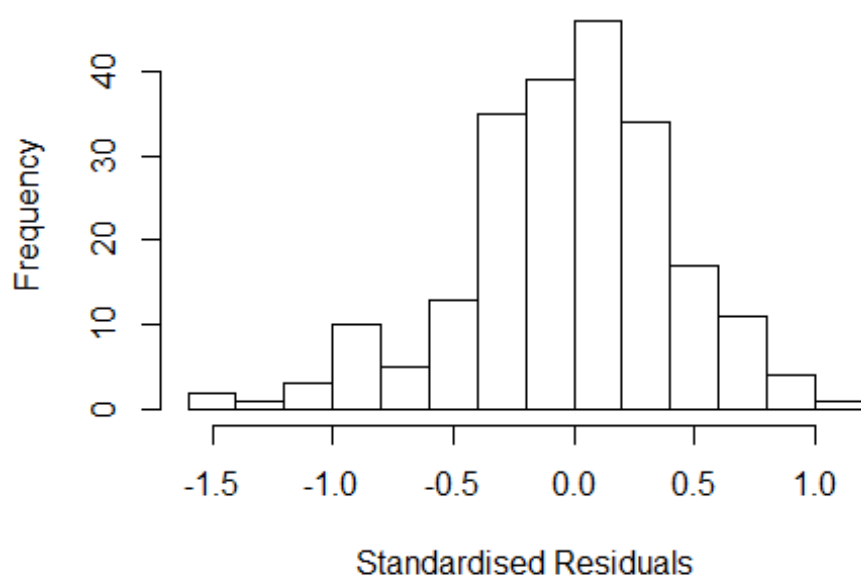
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5407 -0.2506 0.0248 0.2781 1.3808
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1471 0.8595 1.33 0.1835
## FirstAuthorFemale1 0.0838 0.0646 1.30 0.1962
## LastAuthorFemale1 0.0251 0.0651 0.39 0.7002
## UniqueAuthors2 0.3731 0.1363 2.74 0.0068 **
## UniqueAuthors3 0.4152 0.1344 3.09 0.0023 **
## UniqueAuthors4 0.4156 0.1499 2.77 0.0061 **
## UniqueAuthors5 0.5268 0.1302 4.05 7.5e-05 ***
## Year1997 -0.2515 0.8272 -0.30 0.7614
## Year1998 -0.4585 0.9784 -0.47 0.6399
## Year1999 -0.0232 0.8125 -0.03 0.9773
```

```

## Year2000          -0.1314      0.8364   -0.16   0.8754
## Year2001          -0.4897      0.8189   -0.60   0.5506
## Year2002          -0.4804      0.8124   -0.59   0.5550
## Year2003          -0.1474      0.8329   -0.18   0.8597
## Year2004          -0.1300      0.8450   -0.15   0.8779
## Year2005          -0.3186      0.8202   -0.39   0.6981
## Year2006          -0.5296      0.8335   -0.64   0.5259
## Year2007          -0.1304      0.8183   -0.16   0.8735
## Year2008          -0.1703      0.8041   -0.21   0.8325
## Year2009          -0.5366      0.8068   -0.67   0.5068
## Year2010          -0.3725      0.8143   -0.46   0.6478
## Year2011          -0.4378      0.8098   -0.54   0.5894
## Year2012          -0.5055      0.8088   -0.63   0.5326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.195, Adjusted R-squared:  0.106
## Convergence in 33 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.102  0.859  0.949  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      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 1.661 1      1.289
## LastAuthorFemale  1.650 1      1.285
## Year              2.609 16      1.030

```

Residuals from first and last author



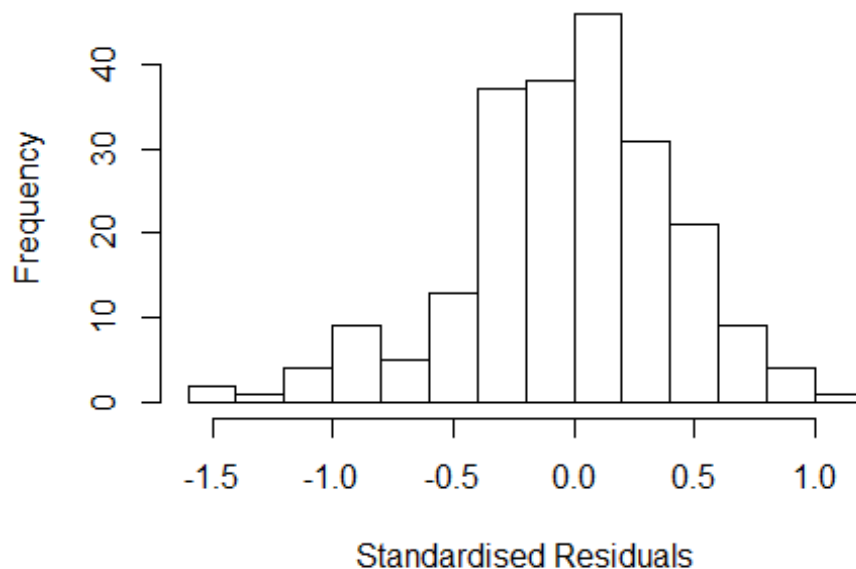
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.47347 -0.27276 0.00682 0.29700 1.01430
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4735 0.4083 3.61 0.00039 ***
## FirstAuthorFemale1 0.0615 0.0684 0.90 0.36965
## LastAuthorFemale1 0.0391 0.0675 0.58 0.56269
## Year1997 -0.2557 0.4257 -0.60 0.54874
## Year1998 -0.5691 0.4690 -1.21 0.22633
## Year1999 0.0165 0.4597 0.04 0.97145
## Year2000 -0.3130 0.4157 -0.75 0.45246
## Year2001 -0.5128 0.4193 -1.22 0.22276
## Year2002 -0.3165 0.4094 -0.77 0.44035
## Year2003 -0.2269 0.4510 -0.50 0.61542
## Year2004 -0.1069 0.4432 -0.24 0.80968
## Year2005 -0.2175 0.4157 -0.52 0.60139
```

```

## Year2006          -0.4236      0.4494   -0.94   0.34708
## Year2007          -0.1502      0.4261   -0.35   0.72485
## Year2008          -0.1012      0.4077   -0.25   0.80424
## Year2009          -0.4938      0.4067   -1.21   0.22611
## Year2010          -0.2948      0.4116   -0.72   0.47461
## Year2011          -0.3229      0.4088   -0.79   0.43051
## Year2012          -0.3764      0.4026   -0.93   0.35093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.0894, Adjusted R-squared:  0.00826
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 203 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.192  0.863   0.951   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 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.621 1      1.273
## Year              1.621 16      1.015

```

Residuals from first author



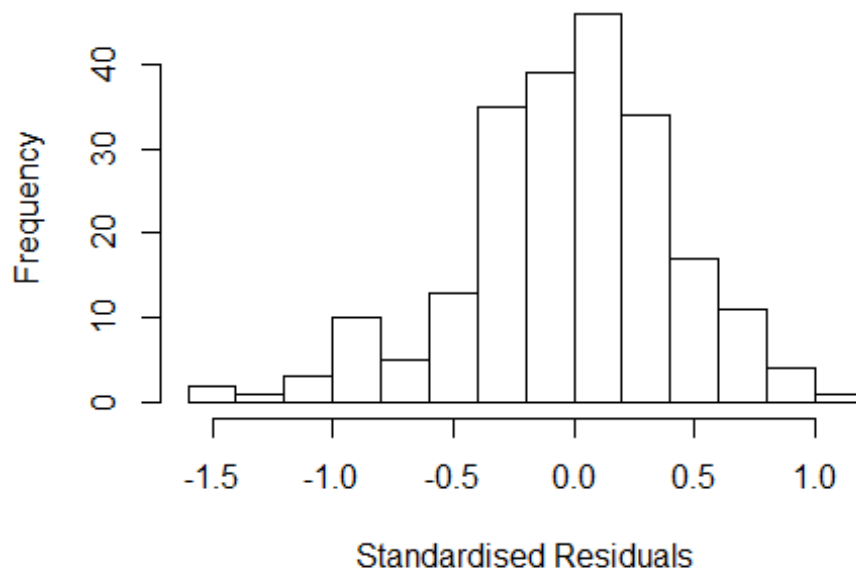
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.486 -0.277 0.008 0.297 1.000
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4857 0.4169 3.56 0.00046 ***
## FirstAuthorFemale1 0.0708 0.0686 1.03 0.30365
## Year1997 -0.2477 0.4332 -0.57 0.56806
## Year1998 -0.5515 0.4783 -1.15 0.25026
## Year1999 0.0238 0.4683 0.05 0.95957
## Year2000 -0.3195 0.4269 -0.75 0.45505
## Year2001 -0.5171 0.4301 -1.20 0.23070
## Year2002 -0.3224 0.4202 -0.77 0.44383
## Year2003 -0.2111 0.4609 -0.46 0.64738
## Year2004 -0.1163 0.4498 -0.26 0.79626
## Year2005 -0.2324 0.4279 -0.54 0.58755
## Year2006 -0.4127 0.4780 -0.86 0.38890
```

```

## Year2007          -0.1412      0.4370   -0.32  0.74696
## Year2008          -0.0969      0.4195   -0.23  0.81763
## Year2009          -0.4920      0.4153   -1.18  0.23758
## Year2010          -0.2928      0.4198   -0.70  0.48633
## Year2011          -0.3252      0.4217   -0.77  0.44151
## Year2012          -0.3721      0.4154   -0.90  0.37132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0871, Adjusted R-squared:  0.0107
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.868  0.953  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      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 1.621 1          1.273
## Year              1.621 16          1.015

```

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
## -1.52191 -0.27351 0.00259 0.28971 0.98717
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5219 0.3759 4.05 7.3e-05 ***
## LastAuthorFemale1 0.0526 0.0674 0.78 0.44
## Year1997 -0.2847 0.4011 -0.71 0.48
## Year1998 -0.5695 0.4315 -1.32 0.19
## Year1999 -0.0387 0.4516 -0.09 0.93
## Year2000 -0.3634 0.3835 -0.95 0.34
## Year2001 -0.5370 0.3912 -1.37 0.17
## Year2002 -0.3455 0.3806 -0.91 0.37
## Year2003 -0.2587 0.4224 -0.61 0.54
## Year2004 -0.1264 0.4224 -0.30 0.77
## Year2005 -0.2174 0.3911 -0.56 0.58
## Year2006 -0.4453 0.4244 -1.05 0.30
```

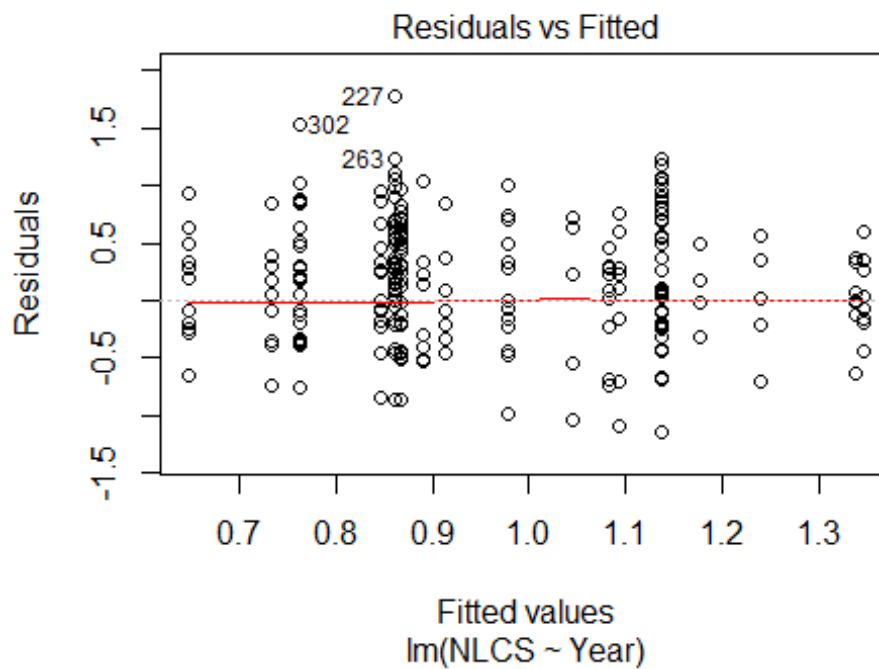


```

## Year2007          -0.1777      0.4020    -0.44      0.66
## Year2008          -0.1246      0.3805    -0.33      0.74
## Year2009          -0.5151      0.3803    -1.35      0.18
## Year2010          -0.3146      0.3854    -0.82      0.42
## Year2011          -0.3320      0.3822    -0.87      0.39
## Year2012          -0.3879      0.3766    -1.03      0.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0864, Adjusted R-squared:  0.0099
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 202 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.153  0.870   0.951   0.895   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      4.52e-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 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
##   11   11   8   13   6   11   15   13   11   18   18   17   23   74   48
## 2011 2012
##   52   53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   10    5   10    5    8    8    8    6   11   15   15   19   61   39
## 2011 2012

```

```
## 46 47
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 10 5 10 5 7 8 8 6 10 13 14 16 60 38
## 2011 2012
## 46 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 = 20, df = 16, p-value = 0.2
```



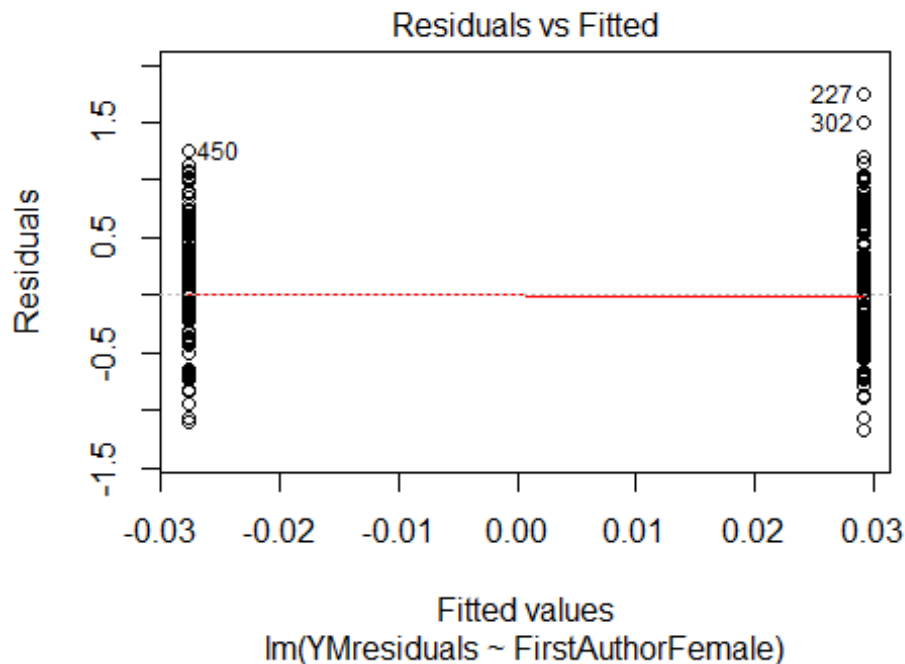
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 2.17650229898678"
## [1] "Male first author team size 2018 geometric mean: 1.50321719249731"

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

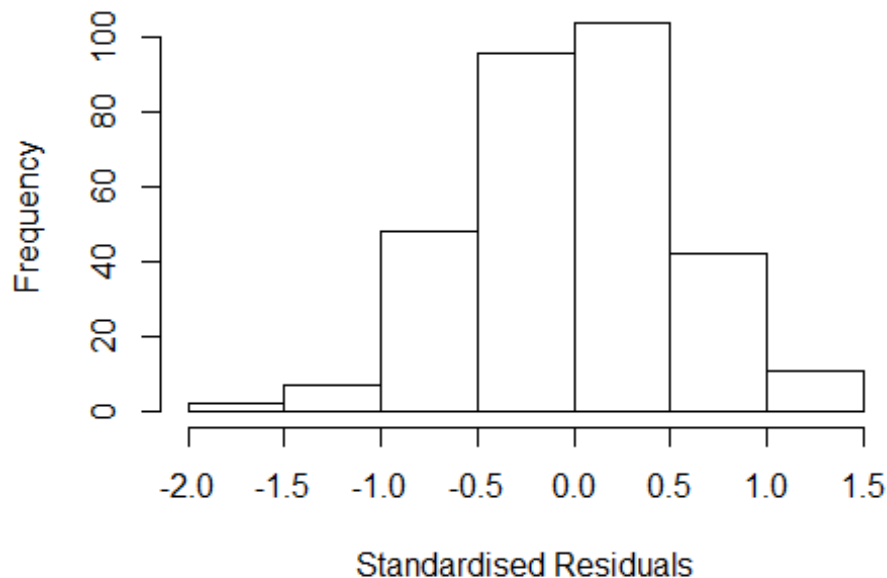
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.20491731338148"
## [1] "Male last author team size 2018 geometric mean: 1.591819444754"

## 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 = 0.1
## 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 3.915  1      1.979
## LastAuthorFemale  4.261  1      2.064
## UniqueAuthors    3.384  4      1.165
## Year              4.134 16      1.045
```

Residuals from first and last author and team size



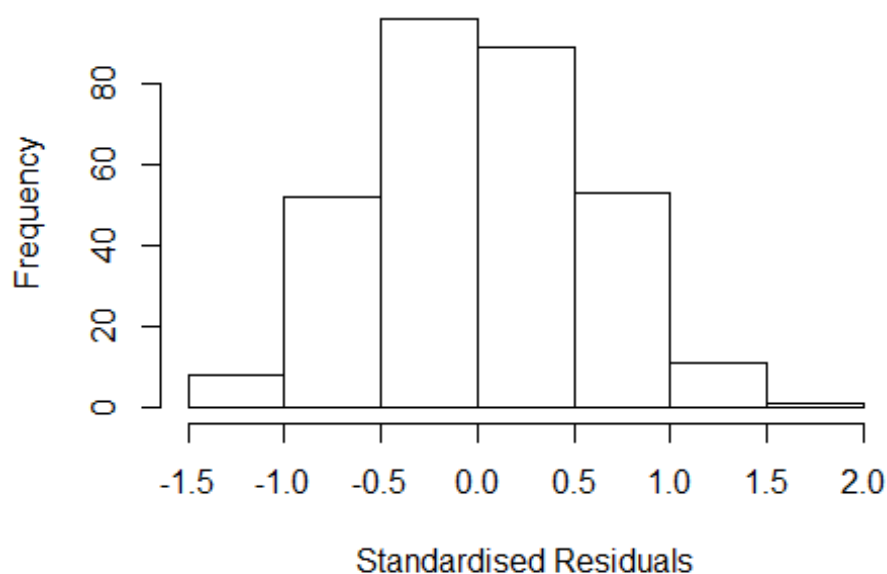
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.53172 -0.37462 0.00724 0.36700 1.38425
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2199 0.1562 7.81 1.1e-13 ***
## FirstAuthorFemale1 0.0643 0.1157 0.56 0.5785
## LastAuthorFemale1 -0.0379 0.1170 -0.32 0.7460
## UniqueAuthors2 0.1975 0.1048 1.88 0.0606 .
## UniqueAuthors3 0.5550 0.1207 4.60 6.5e-06 ***
## UniqueAuthors4 0.4706 0.1374 3.43 0.0007 ***
## UniqueAuthors5 0.2057 0.1564 1.32 0.1895
## Year1997 -0.2147 0.2047 -1.05 0.2951
## Year1998 -0.2559 0.3638 -0.70 0.4824
## Year1999 0.0102 0.1911 0.05 0.9576
```

```

## Year2000          -0.0259      0.2410    -0.11    0.9144
## Year2001          -0.3647      0.2407    -1.52    0.1308
## Year2002          -0.1862      0.2793    -0.67    0.5056
## Year2003          -0.4069      0.2143    -1.90    0.0586 .
## Year2004          -0.3489      0.2084    -1.67    0.0953 .
## Year2005          -0.7652      0.1786    -4.28    2.5e-05 ***
## Year2006          -0.7797      0.1920    -4.06    6.3e-05 ***
## Year2007          -0.4258      0.2266    -1.88    0.0612 .
## Year2008          -0.4367      0.2199    -1.99    0.0479 *
## Year2009          -0.5203      0.1694    -3.07    0.0023 **
## Year2010          -0.5878      0.1774    -3.31    0.0010 **
## Year2011          -0.4437      0.1811    -2.45    0.0149 *
## Year2012          -0.2431      0.1885    -1.29    0.1982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.198, Adjusted R-squared:  0.136
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 277 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.451  0.882  0.952   0.915   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          3.23e-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 4.753 1      2.180
## LastAuthorFemale  4.352 1      2.086
## Year              1.798 16      1.018

```

Residuals from first and last author



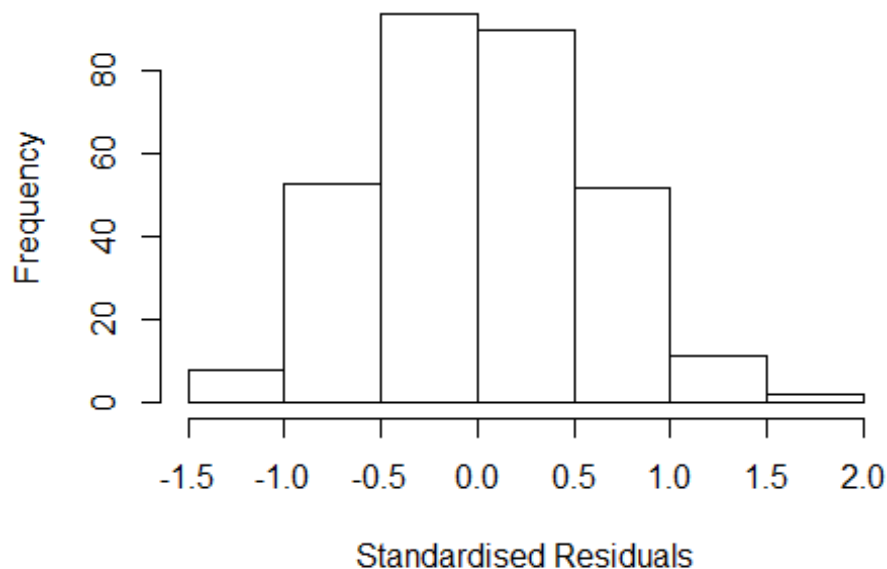
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13327 -0.41212 -0.00676 0.38614 1.68735
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33301 0.11915 11.19 < 2e-16 ***
## FirstAuthorFemale1 0.12653 0.13382 0.95 0.34516
## LastAuthorFemale1 -0.10288 0.12898 -0.80 0.42576
## Year1997 -0.23318 0.17664 -1.32 0.18783
## Year1998 -0.25458 0.39619 -0.64 0.52101
## Year1999 0.00315 0.15182 0.02 0.98344
## Year2000 -0.10072 0.24291 -0.41 0.67870
## Year2001 -0.35916 0.20088 -1.79 0.07483 .
## Year2002 -0.24000 0.26164 -0.92 0.35974
## Year2003 -0.48497 0.21336 -2.27 0.02376 *
## Year2004 -0.17143 0.17541 -0.98 0.32923
## Year2005 -0.69526 0.16593 -4.19 3.7e-05 ***
```

```

## Year2006      -0.81169      0.16623      -4.88      1.7e-06 ***
## Year2007      -0.36226      0.21375      -1.69      0.09119 .
## Year2008      -0.44363      0.18749      -2.37      0.01863 *
## Year2009      -0.51289      0.14505      -3.54      0.00047 ***
## Year2010      -0.62401      0.15108      -4.13      4.7e-05 ***
## Year2011      -0.46404      0.15664      -2.96      0.00330 **
## Year2012      -0.22339      0.16651      -1.34      0.18077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.109, Adjusted R-squared:  0.0534
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 288 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.397  0.874  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      3.23e-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 1.452 1      1.205
## Year      1.452 16      1.012

```

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
## -1.153332 -0.393450 -0.000265 0.395023 1.786146
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32380 0.11994 11.04 < 2e-16 ***
## FirstAuthorFemale1 0.04640 0.07478 0.62 0.53540
## Year1997 -0.24471 0.17224 -1.42 0.15646
## Year1998 -0.25914 0.39404 -0.66 0.51128
## Year1999 0.00824 0.15188 0.05 0.95678
## Year2000 -0.10690 0.24241 -0.44 0.65957
## Year2001 -0.38741 0.19604 -1.98 0.04907 *
## Year2002 -0.23459 0.26274 -0.89 0.37267
## Year2003 -0.48025 0.22177 -2.17 0.03116 *
## Year2004 -0.18068 0.17014 -1.06 0.28916
## Year2005 -0.70022 0.16591 -4.22 3.3e-05 ***
## Year2006 -0.82388 0.16528 -4.98 1.1e-06 ***
```

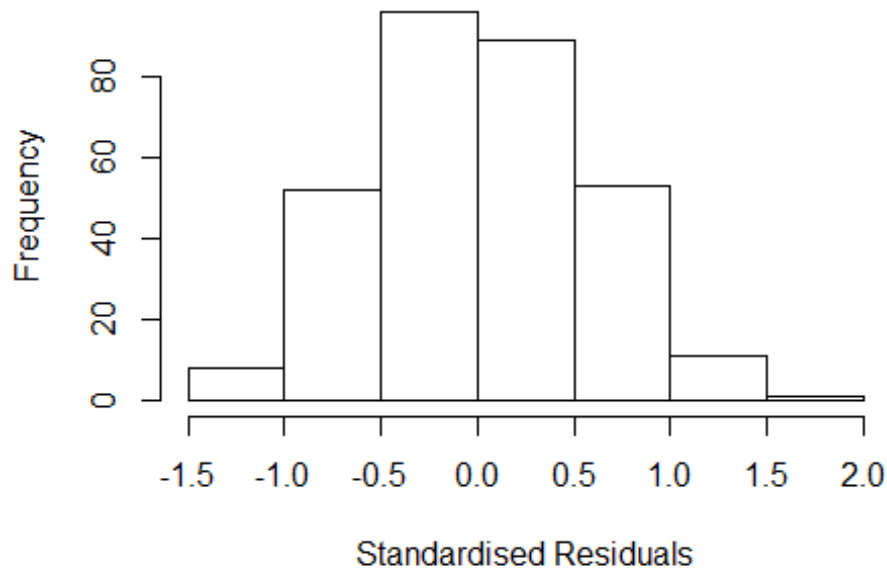


```

## Year2007          -0.35359      0.21461      -1.65      0.10052
## Year2008          -0.43224      0.18442      -2.34      0.01976 *
## Year2009          -0.52235      0.14417      -3.62      0.00034 ***
## Year2010          -0.62543      0.15117      -4.14      4.6e-05 ***
## Year2011          -0.46517      0.15703      -2.96      0.00330 **
## Year2012          -0.21688      0.16756      -1.29      0.19658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.598
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0537
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 285 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.352  0.879   0.953   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      3.23e-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 1.315 1      1.147
## Year              1.315 16      1.009

```

Residuals from last author



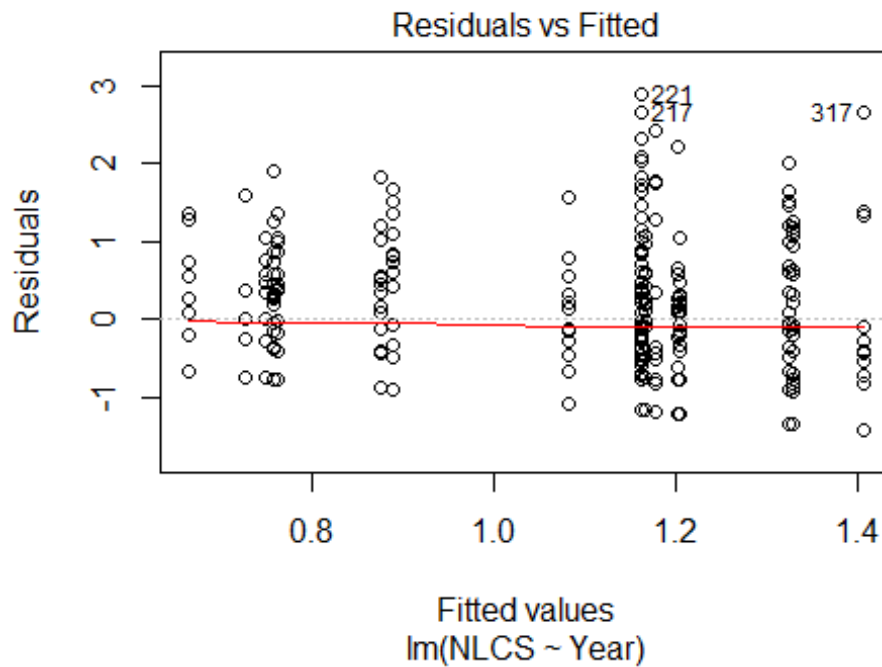
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13811 -0.41153 -0.00751 0.38592 1.81349
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.34552 0.11834 11.37 < 2e-16 ***
## LastAuthorFemale1 -0.00852 0.07114 -0.12 0.9048
## Year1997 -0.24264 0.17611 -1.38 0.1693
## Year1998 -0.25232 0.39279 -0.64 0.5211
## Year1999 -0.00323 0.15189 -0.02 0.9830
## Year2000 -0.09209 0.24191 -0.38 0.7037
## Year2001 -0.39388 0.19460 -2.02 0.0439 *
## Year2002 -0.21125 0.25618 -0.82 0.4103
## Year2003 -0.47761 0.21608 -2.21 0.0279 *
## Year2004 -0.16623 0.16887 -0.98 0.3258
## Year2005 -0.68803 0.16540 -4.16 4.2e-05 ***
## Year2006 -0.81139 0.16177 -5.02 9.2e-07 ***
```

```

## Year2007          -0.34022      0.20982    -1.62    0.1060
## Year2008          -0.42494      0.18203    -2.33    0.0203 *
## Year2009          -0.52500      0.14342    -3.66    0.0003 ***
## Year2010          -0.61972      0.15041    -4.12    4.9e-05 ***
## Year2011          -0.46209      0.15511    -2.98    0.0031 **
## Year2012          -0.20741      0.16751    -1.24    0.2167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0516
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 292 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.352  0.877  0.957  0.921  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.23e-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: 310"
## [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
##   37   43   30   47   48   62   20   14   20   18   27   29   20   27   31
## 2011 2012
##   20   20
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   23   14   15   28   54   14   11   12   17   24   22   20   24   29
## 2011 2012

```

```
## 19 18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 22 14 13 28 51 12 10 11 14 23 21 17 24 28
## 2011 2012
## 19 17
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 40, df = 16, p-value = 7e-04
```



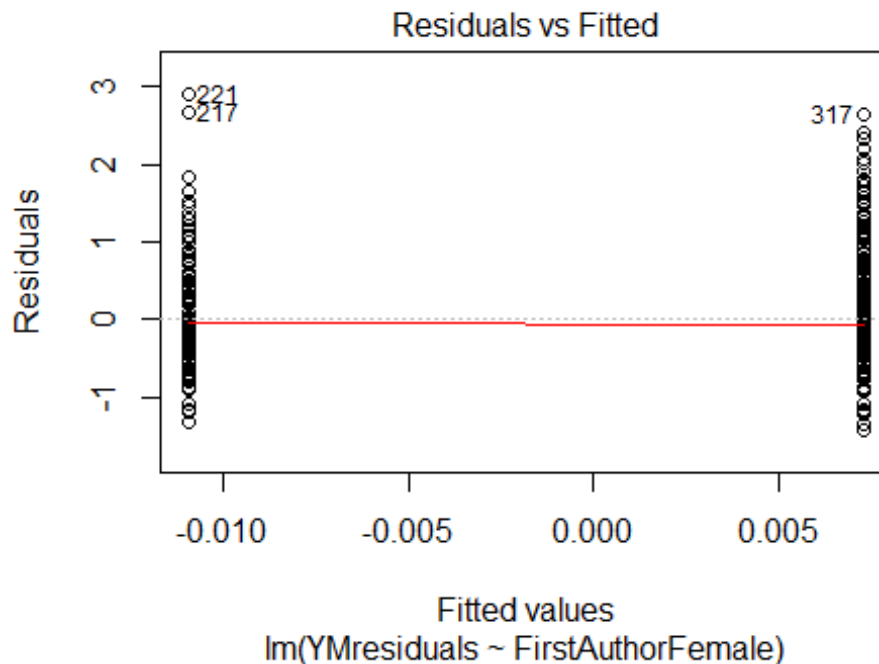
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 3.91746803188132"
## [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
```

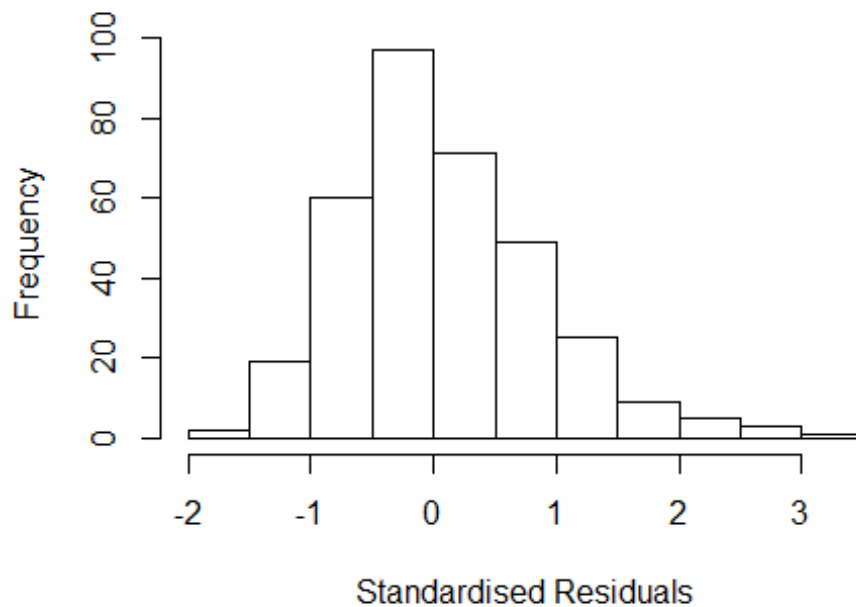
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.91005009931009"
## [1] "Male last author team size 2018 geometric mean: 2.63214802590498"

## 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 = 0.3
## 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 1.827  1      1.351
## LastAuthorFemale  1.990  1      1.411
## UniqueAuthors    3.782  4      1.181
## Year              5.436 16      1.054
```

Residuals from first and last author and team size



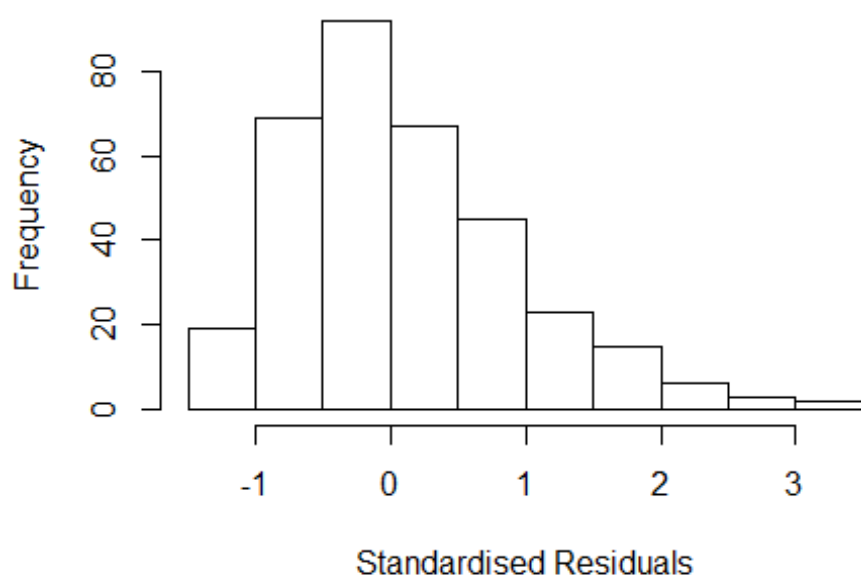
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 221 0035710837 4.055 2001    2911      1    2.749
## 238 0035709407 3.247 2001    2911      1    2.591
## 286 0036314990 3.593 2002    2911      1    3.056
## 306 0037364395 3.424 2003    2911      1    2.652
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8233 -0.4842 -0.0975  0.5398  3.0561
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.66305    0.16151   4.11 5.1e-05 ***
## FirstAuthorFemale1 -0.00831    0.11015  -0.08  0.93992
## LastAuthorFemale1 -0.00495    0.11572  -0.04  0.96594
## UniqueAuthors2     0.10199    0.12077   0.84  0.39902
## UniqueAuthors3     0.55547    0.14803   3.75  0.00021 ***
## UniqueAuthors4     0.64211    0.21271   3.02  0.00274 **
## UniqueAuthors5     0.92368    0.29625   3.12  0.00199 **
```

```

## Year1997          0.07810      0.20557      0.38  0.70427
## Year1998          -0.06544      0.21662     -0.30  0.76278
## Year1999          -0.22629      0.21337     -1.06  0.28968
## Year2000           0.37347      0.31985      1.17  0.24383
## Year2001           0.00593      0.21021      0.03  0.97752
## Year2002          -0.11285      0.38282     -0.29  0.76834
## Year2003           0.12184      0.33862      0.36  0.71922
## Year2004           0.38869      0.35254      1.10  0.27107
## Year2005           0.08740      0.27905      0.31  0.75434
## Year2006          -0.26858      0.21143     -1.27  0.20491
## Year2007           0.30830      0.26075      1.18  0.23794
## Year2008          -0.17831      0.22830     -0.78  0.43537
## Year2009          -0.09046      0.21866     -0.41  0.67938
## Year2010           0.22810      0.21219      1.07  0.28321
## Year2011           0.24980      0.23732      1.05  0.29333
## Year2012           0.16106      0.21912      0.74  0.46284
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.74
## Multiple R-squared:  0.192, Adjusted R-squared:  0.136
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 326 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0501 0.8800 0.9530 0.8990 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      2.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.699 1      1.304
## LastAuthorFemale  1.862 1      1.364
## Year              1.602 16      1.015

```

Residuals from first and last author



```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 216 0035709255 3.493 2001      2911      1      2.600
## 217 0035709316 3.820 2001      2911      1      3.001
## 221 0035710837 4.055 2001      2911      1      3.236
## 286 0036314990 3.593 2002      2911      1      2.634
## 317 0842263623 4.072 2004      2911      1      2.894
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + 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 -0.5287 -0.0681  0.5749  3.2355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.78103    0.15831   4.93 1.3e-06 ***
## FirstAuthorFemale1 -0.00632    0.11311  -0.06  0.956
## LastAuthorFemale1 -0.08001    0.11740  -0.68  0.496
## Year1997         0.04311    0.20565   0.21  0.834
## Year1998        -0.07756    0.20500  -0.38  0.705
## Year1999        -0.30448    0.20781  -1.47  0.144
## Year2000         0.50551    0.30333   1.67  0.097 .
## Year2001         0.11847    0.20859   0.57  0.570
```

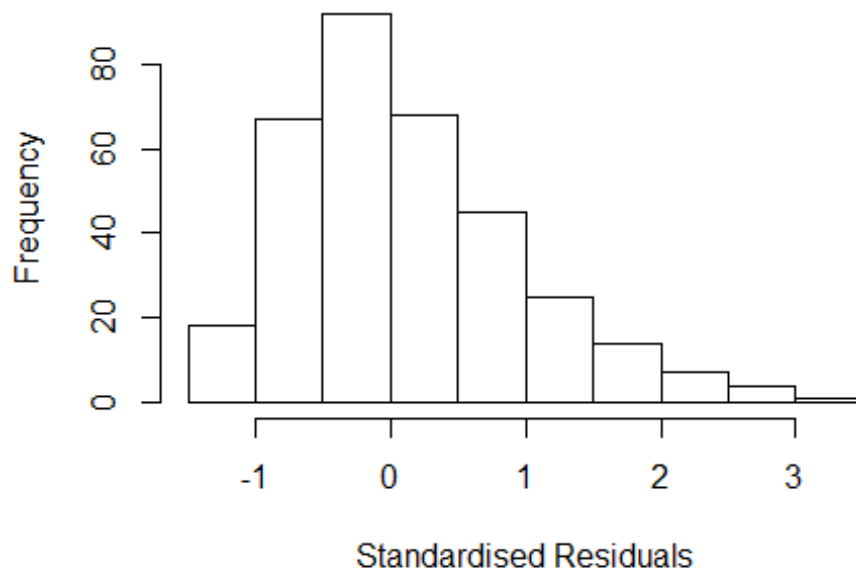


```

## Year2002          0.26440      0.56316      0.47      0.639
## Year2003          0.24367      0.32816      0.74      0.458
## Year2004          0.40315      0.34961      1.15      0.250
## Year2005          0.30640      0.27687      1.11      0.269
## Year2006          0.00775      0.20987      0.04      0.971
## Year2007          0.53783      0.27646      1.95      0.053 .
## Year2008         -0.05277      0.25189     -0.21      0.834
## Year2009          0.10696      0.25558      0.42      0.676
## Year2010          0.39170      0.21830      1.79      0.074 .
## Year2011          0.45999      0.22221      2.07      0.039 *
## Year2012          0.49421      0.19748      2.50      0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.762
## Multiple R-squared:  0.077, Adjusted R-squared:  0.0254
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 315 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0315 0.8700 0.9400 0.8890 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      2.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.286 1          1.134
## Year              1.286 16          1.008

```

Residuals from first author



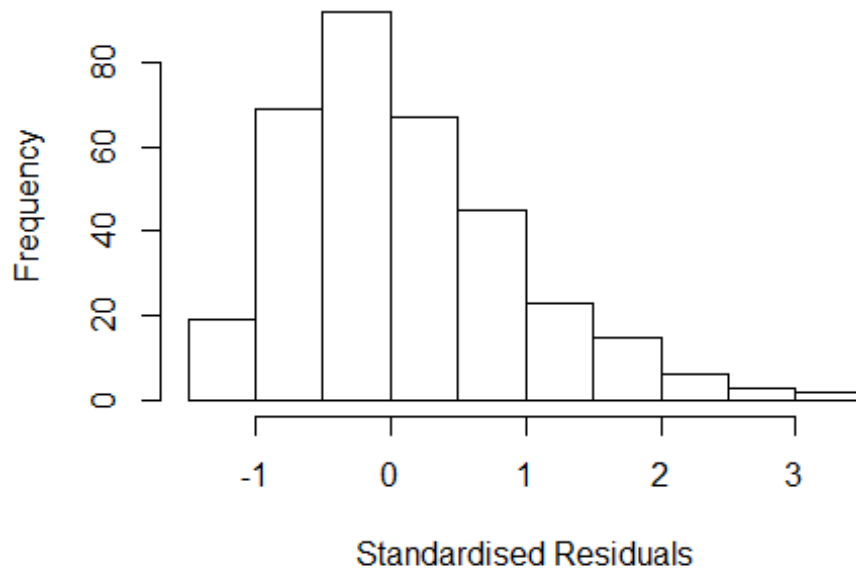
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 216 0035709255 3.493 2001      2911      1      2.600
## 217 0035709316 3.820 2001      2911      1      3.001
## 221 0035710837 4.055 2001      2911      1      3.236
## 286 0036314990 3.593 2002      2911      1      2.634
## 317 0842263623 4.072 2004      2911      1      2.894
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2920 -0.4971 -0.0678  0.5964  3.1868
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76052    0.15534   4.90 1.6e-06 ***
## FirstAuthorFemale1 -0.04871    0.09909  -0.49  0.623
## Year1997         0.04578    0.20628   0.22  0.825
## Year1998        -0.06165    0.20118  -0.31  0.759
## Year1999        -0.29663    0.20811  -1.43  0.155
## Year2000         0.51482    0.31179   1.65  0.100 .
## Year2001         0.10773    0.20755   0.52  0.604
## Year2002         0.21356    0.55313   0.39  0.700
```

```

## Year2003      0.24016    0.32100    0.75    0.455
## Year2004      0.39946    0.33669    1.19    0.236
## Year2005      0.30201    0.27800    1.09    0.278
## Year2006      0.00537    0.21076    0.03    0.980
## Year2007      0.53151    0.27767    1.91    0.056 .
## Year2008     -0.06884    0.25122   -0.27    0.784
## Year2009      0.09203    0.25582    0.36    0.719
## Year2010      0.38059    0.21786    1.75    0.082 .
## Year2011      0.45376    0.22401    2.03    0.044 *
## Year2012      0.49566    0.19706    2.52    0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.742
## Multiple R-squared:  0.0771, Adjusted R-squared:  0.0285
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~ = 1. The remaining 314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0253 0.8700 0.9360 0.8830 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      2.93e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.435 1          1.198
## Year            1.435 16          1.011

```

Residuals from last author



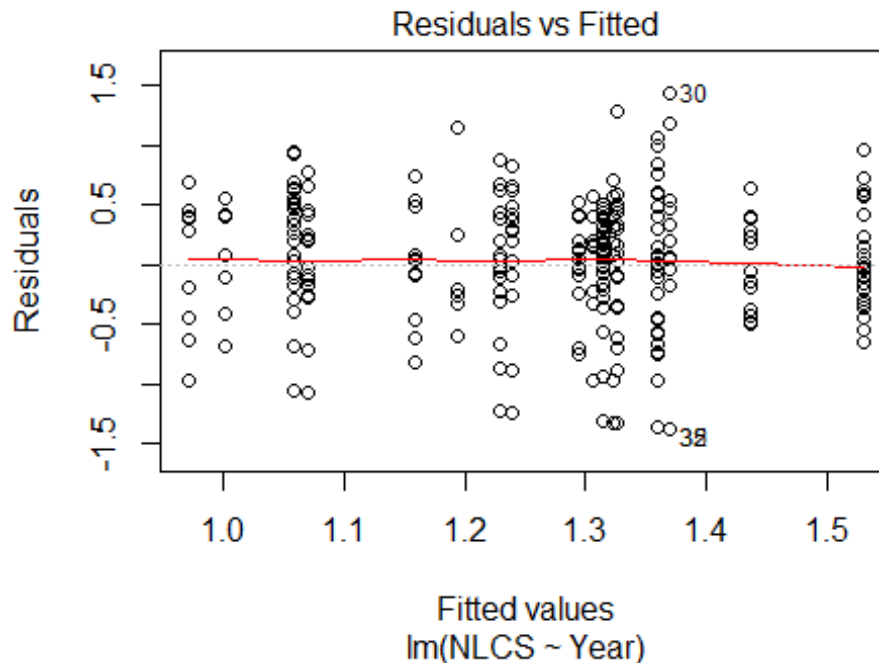
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 216 0035709255 3.493 2001    2911      1    2.600
## 217 0035709316 3.820 2001    2911      1    3.001
## 221 0035710837 4.055 2001    2911      1    3.236
## 286 0036314990 3.593 2002    2911      1    2.634
## 317 0842263623 4.072 2004    2911      1    2.894
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3171 -0.5271 -0.0662  0.5749  3.2411
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.78002    0.15843   4.92 1.4e-06 ***
## LastAuthorFemale1 -0.08324    0.10290  -0.81  0.419
## Year1997         0.04217    0.20405   0.21  0.836
## Year1998        -0.07767    0.20509  -0.38  0.705
## Year1999        -0.30450    0.20780  -1.47  0.144
## Year2000         0.50396    0.30533   1.65  0.100 .
## Year2001         0.11710    0.20846   0.56  0.575
## Year2002         0.26227    0.56387   0.47  0.642
```

```

## Year2003      0.24290      0.32776      0.74      0.459
## Year2004      0.40079      0.35004      1.14      0.253
## Year2005      0.30395      0.27233      1.12      0.265
## Year2006      0.00665      0.20810      0.03      0.975
## Year2007      0.53712      0.27582      1.95      0.052 .
## Year2008     -0.05320      0.25194     -0.21      0.833
## Year2009      0.10489      0.25419      0.41      0.680
## Year2010      0.39110      0.21824      1.79      0.074 .
## Year2011      0.45934      0.22120      2.08      0.039 *
## Year2012      0.49256      0.19378      2.54      0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.76
## Multiple R-squared:  0.0771, Adjusted R-squared:  0.0285
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~ = 1. The remaining 315 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0296 0.8690 0.9390 0.8880 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           2.93e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 341"
## [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
##    7   15   16    8   12   13   13   19   23   18   15   21   33   32   38
## 2011 2012
##   33   21

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    9   13    8    8   11   13   18   14   17   14   18   30   27   35
## 2011 2012
##   31   19
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6    9   13    7    8   11   13   16   14   14   13   17   28   27   35
## 2011 2012
##   31   16
## [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
```



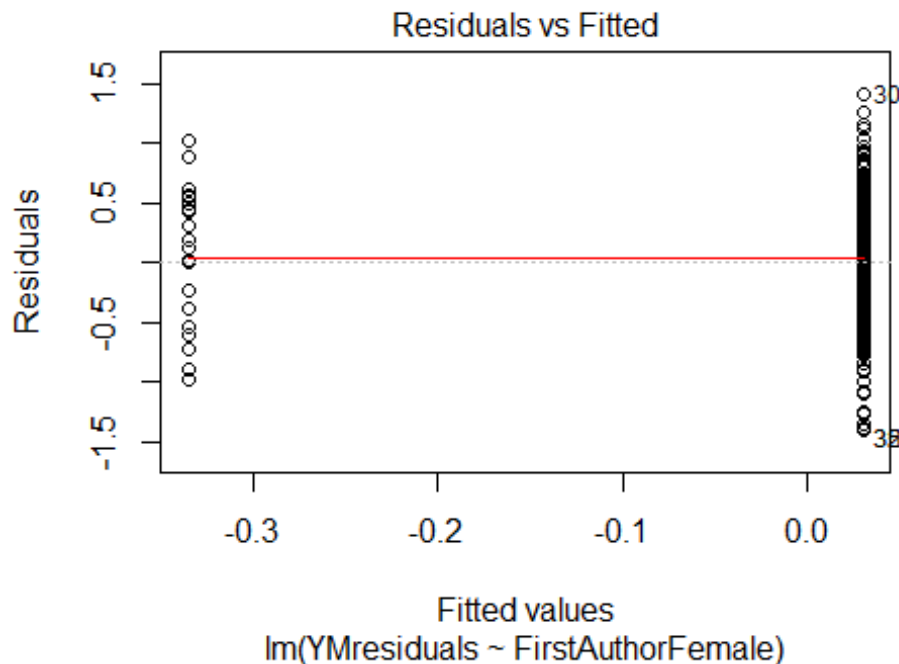
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.67, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 3.18107789337119"
## [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 = 68, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.70928807443503"
## [1] "Male last author team size 2018 geometric mean: 4.16940525104117"

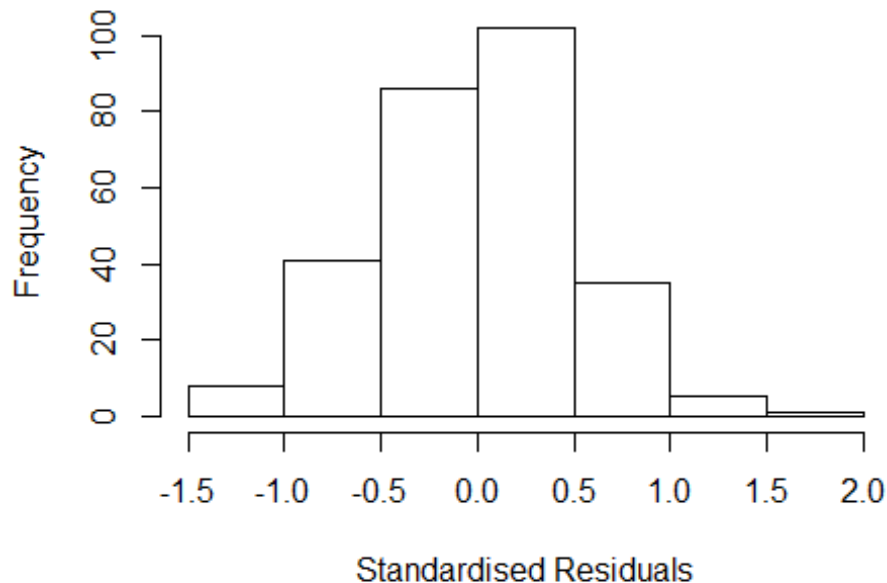
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 70, 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 1.519  1          1.232
```

## LastAuthorFemale	1.460	1	1.208
## UniqueAuthors	4.540	4	1.208
## Year	6.000	16	1.058

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
## -1.4328 -0.3607 0.0338 0.3214 1.5164
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7207 0.2148 3.35 0.00092 ***
## FirstAuthorFemale1 0.2395 0.1325 1.81 0.07182 .
## LastAuthorFemale1 -0.0281 0.0782 -0.36 0.71971
## UniqueAuthors2 0.2427 0.1303 1.86 0.06365 .
## UniqueAuthors3 0.3601 0.1211 2.97 0.00322 **
## UniqueAuthors4 0.4702 0.1235 3.81 0.00018 ***
```

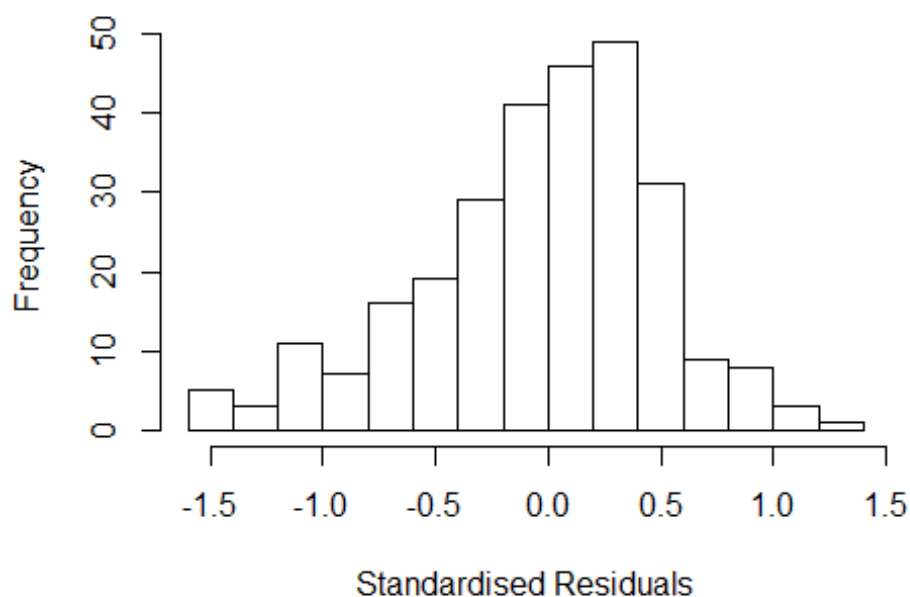


```

## UniqueAuthors5      0.2495      0.1351      1.85  0.06594 .
## Year1997             -0.1998      0.2467     -0.81  0.41871
## Year1998             0.3113      0.3537      0.88  0.37951
## Year1999             0.2477      0.3397      0.73  0.46665
## Year2000             0.2666      0.2385      1.12  0.26462
## Year2001            -0.0770      0.2809     -0.27  0.78430
## Year2002             0.2888      0.2320      1.24  0.21428
## Year2003             0.0524      0.2431      0.22  0.82942
## Year2004            -0.0670      0.2306     -0.29  0.77168
## Year2005            -0.0181      0.2383     -0.08  0.93937
## Year2006             0.1296      0.2184      0.59  0.55337
## Year2007             0.0759      0.2699      0.28  0.77886
## Year2008             0.1633      0.2043      0.80  0.42499
## Year2009             0.2794      0.2171      1.29  0.19925
## Year2010            -0.0460      0.2268     -0.20  0.83944
## Year2011             0.1125      0.2312      0.49  0.62718
## Year2012             0.1635      0.2380      0.69  0.49262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.173, Adjusted R-squared:  0.102
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.329  0.868  0.952  0.903  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      3.60e-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 1.611 1      1.269
## LastAuthorFemale  1.277 1      1.130
## Year              2.027 16      1.022

```

Residuals from first and last author



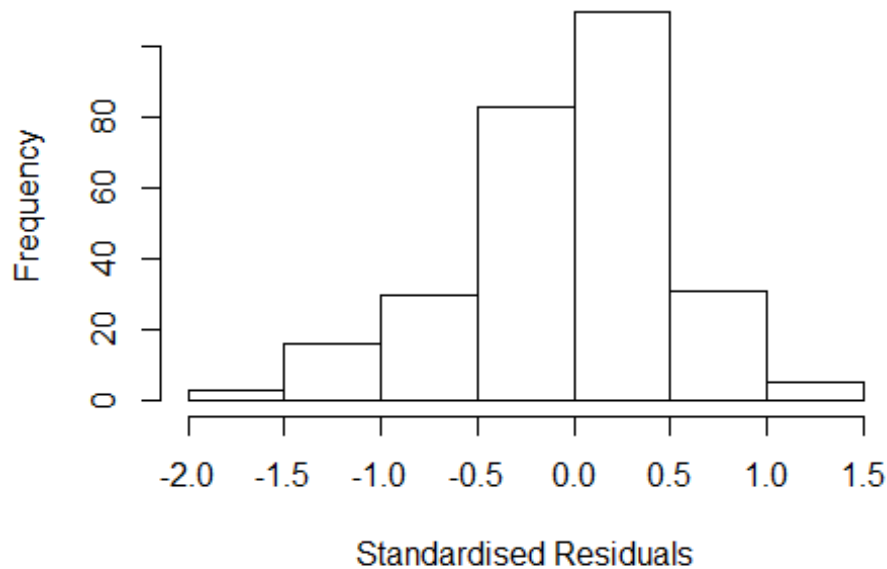
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5266 -0.3270 0.0334 0.3629 1.2386
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9631 0.2253 4.27 2.7e-05 ***
## FirstAuthorFemale1 0.3099 0.1599 1.94 0.054 .
## LastAuthorFemale1 -0.0429 0.0780 -0.55 0.583
## Year1997 -0.2226 0.2966 -0.75 0.454
## Year1998 0.2965 0.3627 0.82 0.414
## Year1999 0.1678 0.3951 0.42 0.671
## Year2000 0.1452 0.2713 0.54 0.593
## Year2001 -0.1806 0.3180 -0.57 0.571
## Year2002 0.1835 0.2657 0.69 0.490
## Year2003 -0.0140 0.2588 -0.05 0.957
## Year2004 -0.0723 0.2621 -0.28 0.783
## Year2005 -0.1313 0.2552 -0.51 0.607
```

```

## Year2006          0.0741      0.2602      0.28      0.776
## Year2007          0.0832      0.2994      0.28      0.781
## Year2008          0.1762      0.2411      0.73      0.465
## Year2009          0.2968      0.2519      1.18      0.240
## Year2010         -0.0766      0.2741     -0.28      0.780
## Year2011          0.1436      0.2638      0.54      0.587
## Year2012          0.2255      0.2727      0.83      0.409
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.106, Adjusted R-squared:  0.044
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 255 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.305  0.869  0.945  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      3.60e-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 1.539 1          1.240
## Year              1.539 16          1.014

```

Residuals from first author



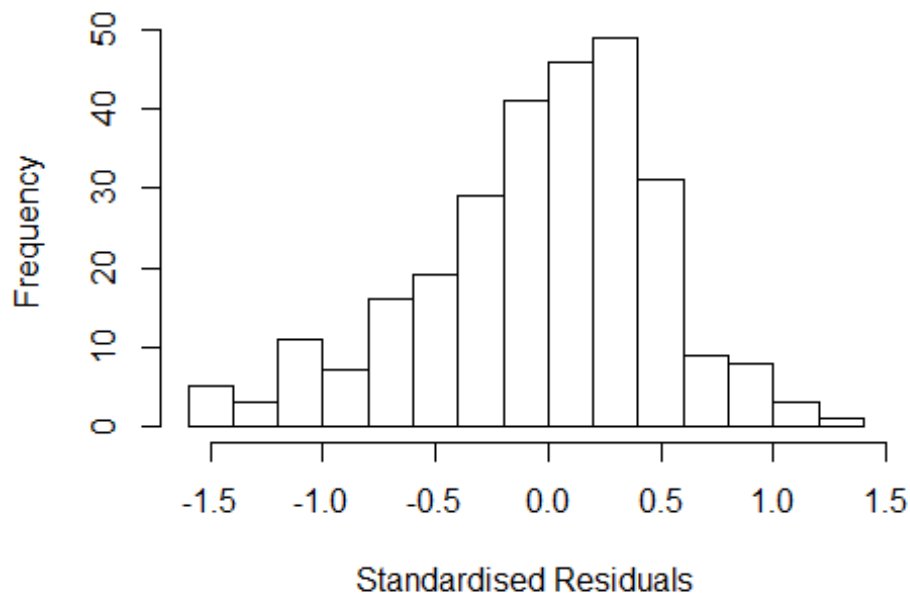
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5424 -0.3342 0.0326 0.3582 1.2656
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9358 0.2125 4.40 1.6e-05 ***
## FirstAuthorFemale1 0.2952 0.1592 1.85 0.065 .
## Year1997 -0.2129 0.3023 -0.70 0.482
## Year1998 0.3114 0.3667 0.85 0.396
## Year1999 0.1839 0.4001 0.46 0.646
## Year2000 0.1558 0.2727 0.57 0.568
## Year2001 -0.1819 0.3243 -0.56 0.575
## Year2002 0.2029 0.2657 0.76 0.446
## Year2003 -0.0149 0.2649 -0.06 0.955
## Year2004 -0.0669 0.2674 -0.25 0.803
## Year2005 -0.1240 0.2585 -0.48 0.632
## Year2006 0.0802 0.2648 0.30 0.762
```

```

## Year2007          0.0952      0.3018      0.32      0.753
## Year2008          0.1832      0.2451      0.75      0.455
## Year2009          0.3072      0.2538      1.21      0.227
## Year2010         -0.0633      0.2754     -0.23      0.818
## Year2011          0.1551      0.2664      0.58      0.561
## Year2012          0.2405      0.2722      0.88      0.378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.106, Adjusted R-squared:  0.048
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.863   0.946   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      3.60e-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 1.271 1      1.127
## Year              1.271 16      1.008

```

Residuals from last author



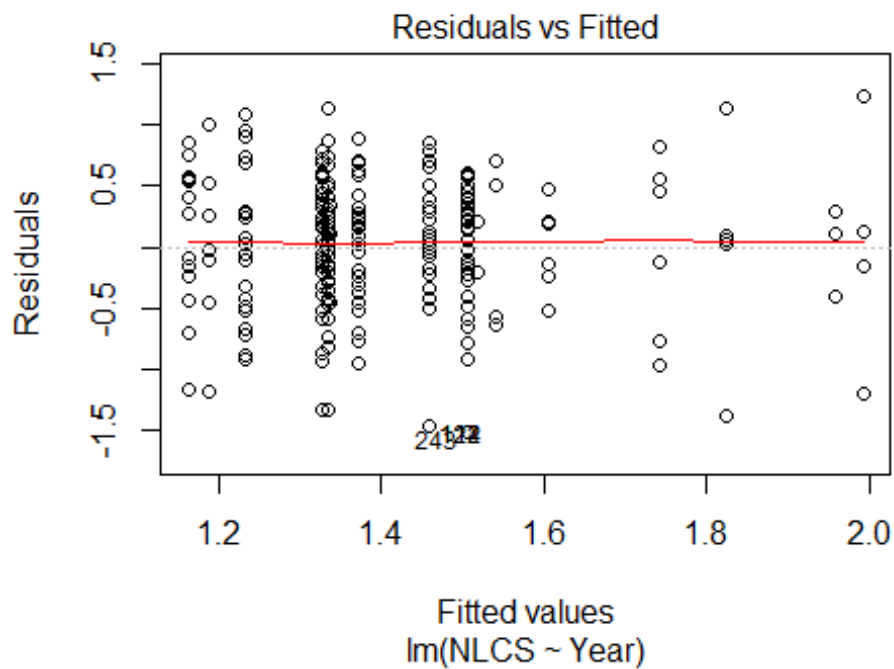
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5371 -0.3429 0.0263 0.3686 1.2583
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1049 0.2348 4.71 4.1e-06 ***
## LastAuthorFemale1 -0.0126 0.0818 -0.15 0.878
## Year1997 -0.0774 0.2795 -0.28 0.782
## Year1998 0.4448 0.3487 1.28 0.203
## Year1999 0.3488 0.3773 0.92 0.356
## Year2000 0.2472 0.2729 0.91 0.366
## Year2001 -0.0916 0.2852 -0.32 0.748
## Year2002 0.3356 0.2441 1.37 0.170
## Year2003 0.0991 0.2470 0.40 0.689
## Year2004 0.0699 0.2402 0.29 0.771
## Year2005 -0.0309 0.2476 -0.12 0.901
## Year2006 0.2167 0.2378 0.91 0.363
```

```

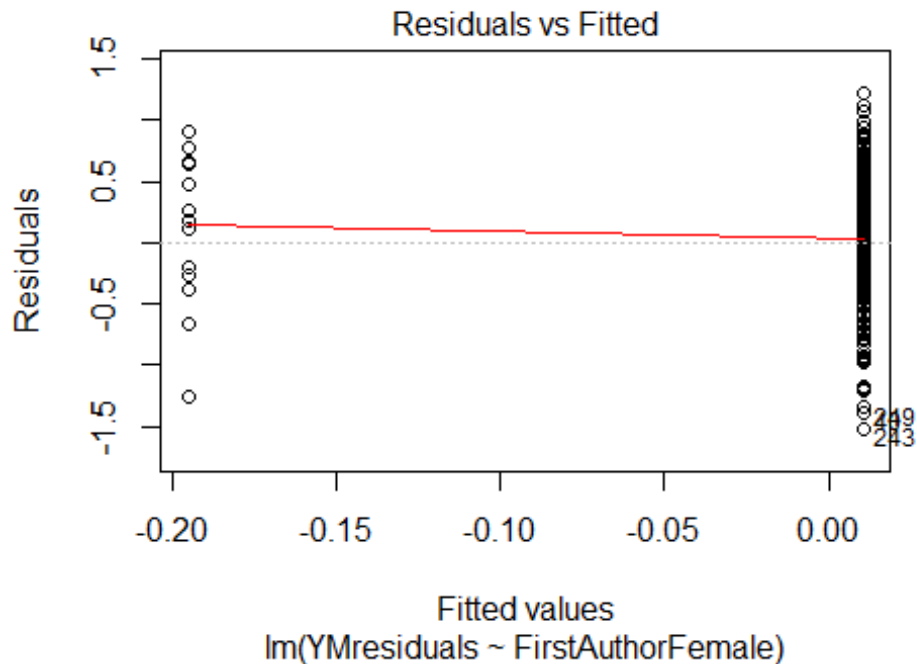
## Year2007          0.2373      0.2781      0.85      0.394
## Year2008          0.2940      0.2230      1.32      0.189
## Year2009          0.4181      0.2309      1.81      0.071 .
## Year2010          0.0342      0.2607      0.13      0.896
## Year2011          0.2786      0.2482      1.12      0.263
## Year2012          0.3478      0.2575      1.35      0.178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.089, Adjusted R-squared:  0.0295
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.296  0.868  0.945  0.888  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      3.60e-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: 278"
## [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
##    1    6    4    6   13   14   11    6    7    5   19   35   27   35   40
## 2011 2012
##   53   51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    2    3    3    6    6    6    4    7    4   14   33   24   33   36
## 2011 2012

```

```
## 48 43
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 0 2 3 3 6 5 6 4 7 4 14 30 22 32 32
## 2011 2012
## 44 40
## [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 = 15, p-value = 0.3
```

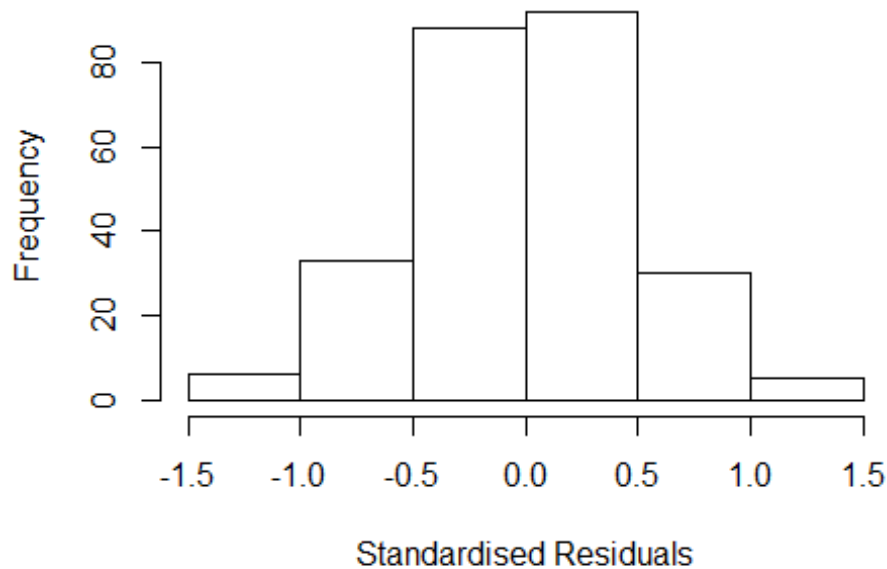


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.1, df = 1, p-value = 0.08
```

```
## [1] "Female first author team size 2018 geometric mean: 3.24672593972156"
## [1] "Male first author team size 2018 geometric mean: 3"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 95, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.13550964279004"
## [1] "Male last author team size 2018 geometric mean: 4.40923831739395"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, 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 1.243 1 1.115
## LastAuthorFemale 1.449 1 1.204
## UniqueAuthors 5.301 4 1.232
## Year 7.168 15 1.068
```

Residuals from first and last author and team size



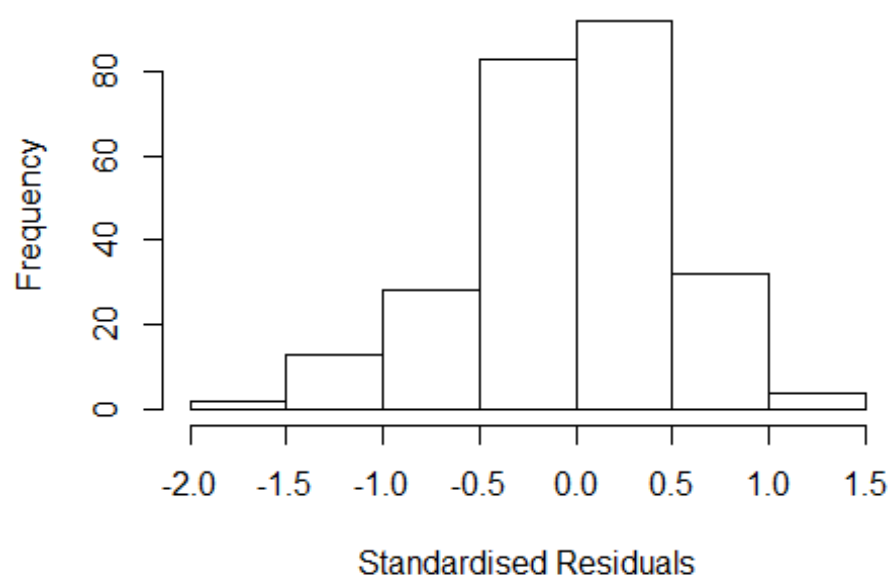
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41742 -0.29596 0.00574 0.30219 1.18788
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0983 0.2616 4.20 3.8e-05 ***
## FirstAuthorFemale1 0.1796 0.1906 0.94 0.34697
## LastAuthorFemale1 0.2406 0.0817 2.94 0.00356 **
## UniqueAuthors2 0.3744 0.1358 2.76 0.00630 **
## UniqueAuthors3 0.4808 0.1288 3.73 0.00024 ***
## UniqueAuthors4 0.6778 0.1448 4.68 4.9e-06 ***
## UniqueAuthors5 0.7583 0.1264 6.00 7.5e-09 ***
## Year1998 0.0628 0.1800 0.35 0.72730
## Year1999 -0.2397 0.3223 -0.74 0.45771
## Year2000 -0.0732 0.2162 -0.34 0.73533
```

```

## Year2001          -0.0305      0.4007   -0.08   0.93946
## Year2002          -0.0290      0.2972   -0.10   0.92225
## Year2003           0.0368      0.7210    0.05   0.95932
## Year2004          -0.6766      0.2880   -2.35   0.01965 *
## Year2005          -0.5338      0.3571   -1.49   0.13632
## Year2006          -0.5883      0.2661   -2.21   0.02800 *
## Year2007          -0.4458      0.2056   -2.17   0.03113 *
## Year2008          -0.6213      0.2160   -2.88   0.00439 **
## Year2009          -0.6046      0.1979   -3.06   0.00251 **
## Year2010          -0.4583      0.2140   -2.14   0.03322 *
## Year2011          -0.7061      0.2052   -3.44   0.00069 ***
## Year2012          -0.6515      0.2036   -3.20   0.00157 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.492
## Multiple R-squared:  0.264, Adjusted R-squared:  0.197
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.387  0.882  0.960  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      3.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.270 1 1.127
## LastAuthorFemale 1.363 1 1.168
## Year 1.552 15 1.015

```

Residuals from first and last author



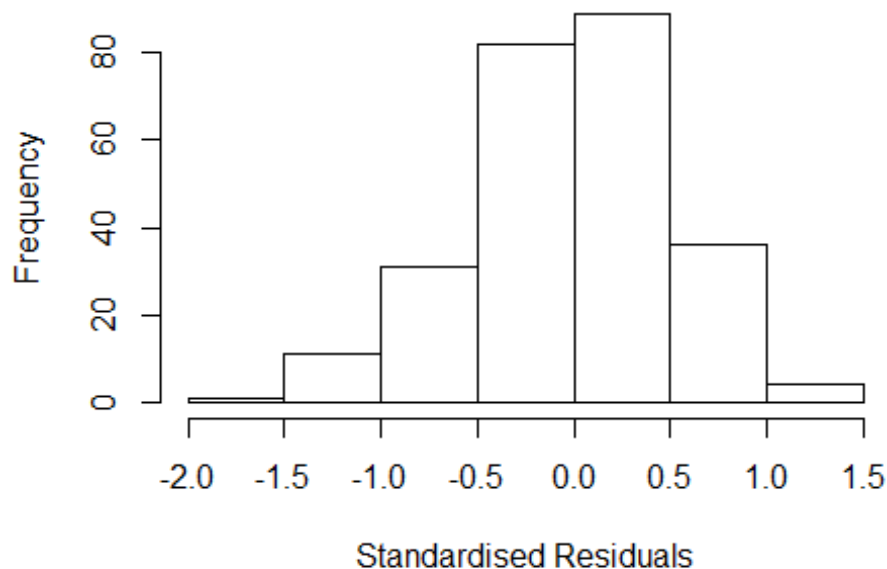
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5451 -0.3171 0.0115 0.3120 1.2602
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2543 0.3618 3.47 0.00063 ***
## FirstAuthorFemale1 0.1167 0.2887 0.40 0.68639
## LastAuthorFemale1 0.1475 0.1003 1.47 0.14281
## Year1998 0.4456 0.2367 1.88 0.06104 .
## Year1999 -0.1304 0.2226 -0.59 0.55849
## Year2000 0.1158 0.2080 0.56 0.57838
## Year2001 0.4268 0.3844 1.11 0.26803
## Year2002 0.3353 0.4373 0.77 0.44394
## Year2003 0.4453 0.6574 0.68 0.49886
## Year2004 -0.2633 0.2939 -0.90 0.37131
## Year2005 0.0562 0.4081 0.14 0.89052
## Year2006 -0.2416 0.2733 -0.88 0.37770
```

```

## Year2007          -0.0380      0.1760   -0.22  0.82935
## Year2008          -0.3064      0.2004   -1.53  0.12755
## Year2009          -0.1063      0.1732   -0.61  0.53997
## Year2010           0.0266      0.1759    0.15  0.88001
## Year2011          -0.1517      0.1683   -0.90  0.36802
## Year2012          -0.1106      0.1844   -0.60  0.54921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0362
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 235 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.321  0.870  0.955  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      3.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.224 1          1.106
## Year              1.224 15          1.007

```

Residuals from first author



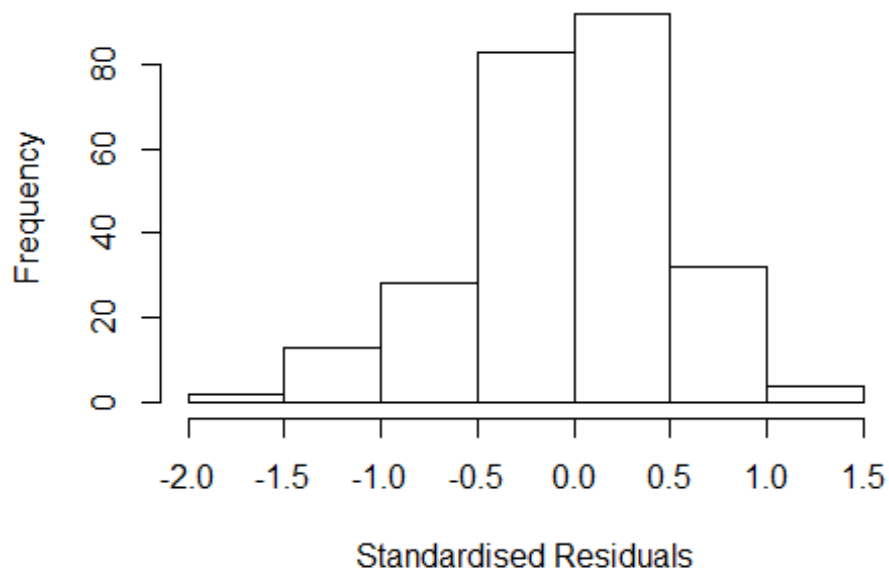
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.52726 -0.31287 0.00952 0.31271 1.25991
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.40088 0.36205 3.87 0.00014 ***
## FirstAuthorFemale1 0.11762 0.32776 0.36 0.72001
## Year1998 0.44557 0.23669 1.88 0.06099 .
## Year1999 -0.17523 0.25539 -0.69 0.49331
## Year2000 0.08826 0.20874 0.42 0.67280
## Year2001 0.42588 0.38481 1.11 0.26954
## Year2002 0.30047 0.44164 0.68 0.49695
## Year2003 0.44559 0.65569 0.68 0.49744
## Year2004 -0.29068 0.30869 -0.94 0.34733
## Year2005 0.01836 0.44665 0.04 0.96724
## Year2006 -0.28728 0.26113 -1.10 0.27238
## Year2007 -0.05593 0.17619 -0.32 0.75119
```

```

## Year2008          -0.30621    0.20029   -1.53  0.12765
## Year2009          -0.11978    0.17407   -0.69  0.49206
## Year2010           0.00876    0.17576    0.05  0.96029
## Year2011          -0.17202    0.16729   -1.03  0.30486
## Year2012          -0.15121    0.17955   -0.84  0.40054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.093, Adjusted R-squared:  0.0318
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 236 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.878  0.958  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      3.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.257 1          1.121
## Year            1.257 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
## -1.54060 -0.31260 0.00405 0.31205 1.25934
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3712 0.1847 7.42 2e-12 ***
## LastAuthorFemale1 0.1473 0.1024 1.44 0.152
## Year1998 0.4455 0.2366 1.88 0.061 .
## Year1999 -0.1306 0.2226 -0.59 0.558
## Year2000 0.1156 0.2079 0.56 0.579
## Year2001 0.4241 0.3855 1.10 0.272
## Year2002 0.3054 0.4053 0.75 0.452
## Year2003 0.4462 0.6524 0.68 0.495
## Year2004 -0.2642 0.2939 -0.90 0.370
## Year2005 0.0562 0.4064 0.14 0.890
## Year2006 -0.2434 0.2724 -0.89 0.372
## Year2007 -0.0396 0.1760 -0.23 0.822
```



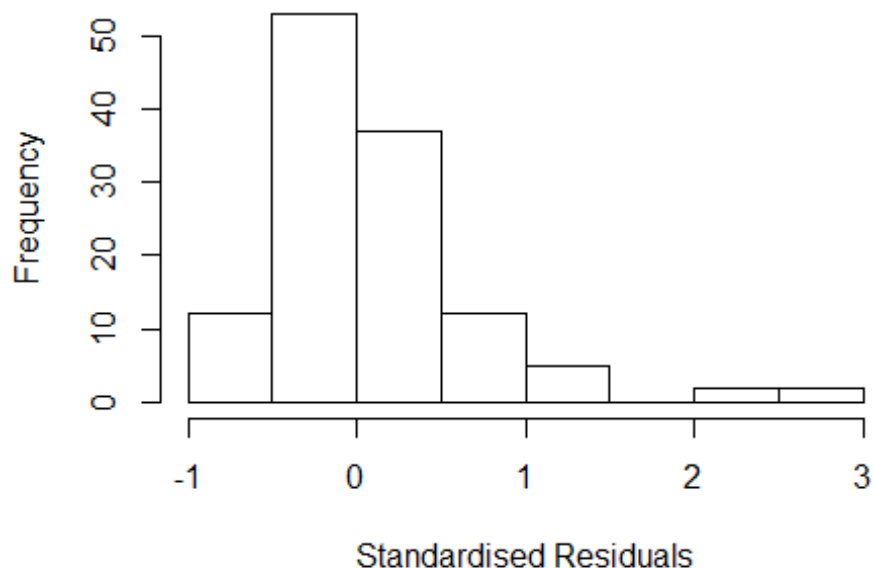
```

## Year2008          -0.3059      0.2002    -1.53      0.128
## Year2009          -0.1064      0.1731    -0.61      0.539
## Year2010           0.0221      0.1758     0.13      0.900
## Year2011          -0.1552      0.1677    -0.93      0.356
## Year2012          -0.1245      0.1830    -0.68      0.497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.0981, Adjusted R-squared:  0.0373
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 233 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.332  0.873   0.954   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      3.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 254"
## [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
##    3    4    4    4    6    2    7    7    4    4    9    7   13   16   27
## 2011 2012
##   24   31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    2    1    0    3    1    4    6    4    4    7    5   12   14   23
## 2011 2012
##   21   23

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    2    1    0    3    1    4    6    4    4    7    5   10   13   21
## 2011 2012
##   17   22
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.72350825419833"
## [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 2.433e+00  1          1.560
## LastAuthorFemale  3.518e+00  1          1.876
## UniqueAuthors    4.073e+15  4          89.379
## Year              1.397e+16 15          3.453
```

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
## 136 79959685959 2.630 2011    2914      1    2.518
## 174 84860317886 3.474 2012    2914      1    2.916
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
```

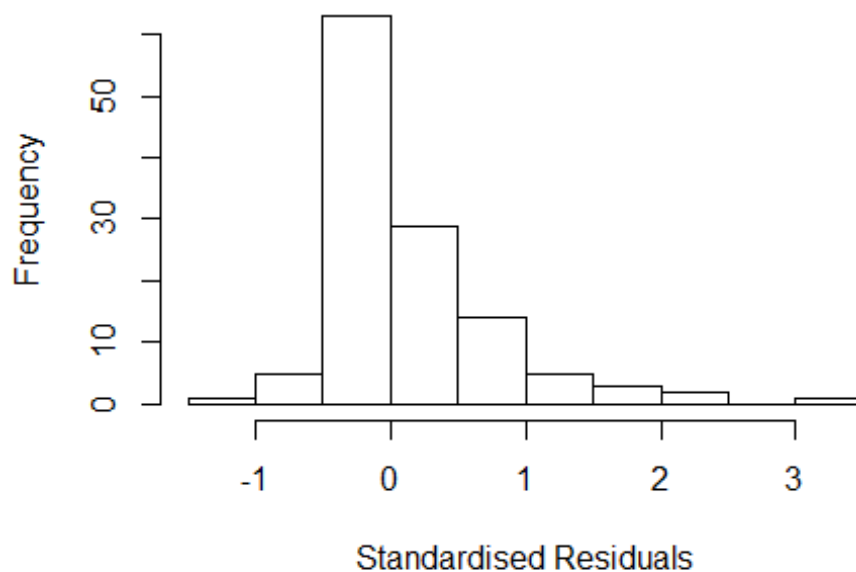
```

##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -0.9238 -0.2368 -0.0917  0.3618  2.9164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.04637    0.21478   0.22  0.8295
## FirstAuthorFemale1 -0.09628    0.17967  -0.54  0.5932
## LastAuthorFemale1  0.22651    0.16383   1.38  0.1698
## UniqueAuthors2     0.33437    0.13985   2.39  0.0187 *
## UniqueAuthors3     0.33576    0.17885   1.88  0.0634 .
## UniqueAuthors4    -0.00153    0.18441  -0.01  0.9934
## UniqueAuthors5     0.59430    0.21913   2.71  0.0079 **
## Year1997          1.58266    0.49444   3.20  0.0018 **
## Year1998           0.67693    0.30188   2.24  0.0271 *
## Year2000           0.77334    0.44851   1.72  0.0877 .
## Year2001          -0.17660    0.23716  -0.74  0.4582
## Year2002           0.99017    0.36739   2.70  0.0082 **
## Year2003           1.32937    0.30327   4.38 2.9e-05 ***
## Year2004           0.71085    0.31629   2.25  0.0268 *
## Year2005           0.51552    0.41162   1.25  0.2133
## Year2006           0.74719    0.42284   1.77  0.0802 .
## Year2007           0.33472    0.37881   0.88  0.3790
## Year2008           0.11575    0.29871   0.39  0.6992
## Year2009          -0.07758    0.25347  -0.31  0.7602
## Year2010           0.04323    0.26018   0.17  0.8684
## Year2011          -0.06349    0.23760  -0.27  0.7899
## Year2012           0.04528    0.25093   0.18  0.8572
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.535, Adjusted R-squared:  0.438
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 3 observations c(68,91,115) are outliers with |weight| = 0 ( < 0.00081);
## 8 weights are ~= 1. The remaining 112 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.022  0.893  0.948  0.898  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      8.13e-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 6.487 1 2.547
## LastAuthorFemale 6.511 1 2.552
## Year 3.927 15 1.047
```

Residuals from first and last author



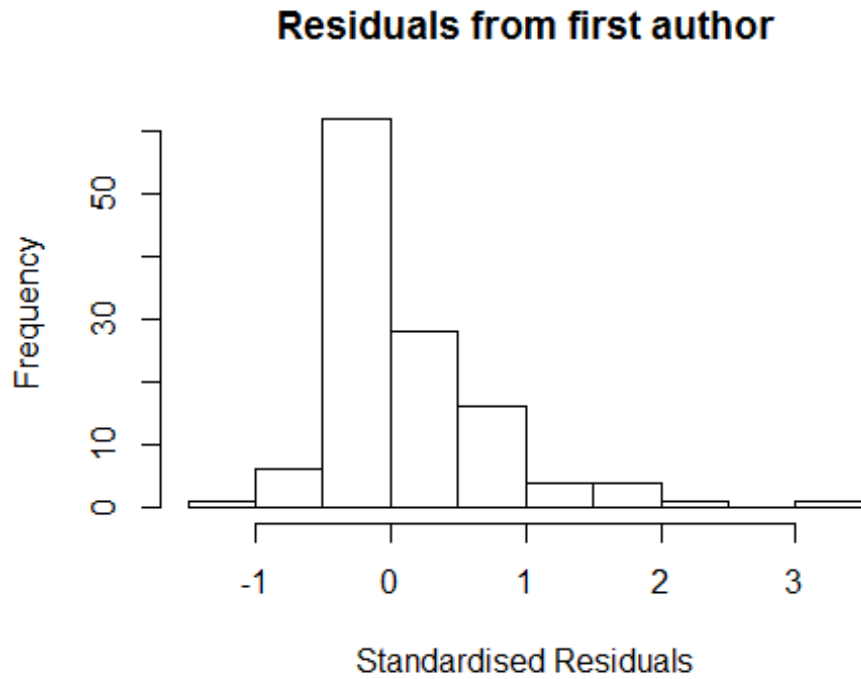
```
## [1] "List of 1 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 174 84860317886 3.474 2012 2914 1 3.128
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.334 -0.331 -0.161 0.451 3.128
```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.18037    0.15830    1.14  0.25712
## FirstAuthorFemale1 -0.12131    0.22428   -0.54  0.58972
## LastAuthorFemale1  0.18089    0.21856    0.83  0.40974
## Year1997         1.51855    0.46644    3.26  0.00152 **
## Year1998         0.61205    0.18983    3.22  0.00168 **
## Year2000         0.70106    0.40495    1.73  0.08635 .
## Year2001        -0.23995    0.18983   -1.26  0.20903
## Year2002         1.22644    0.37438    3.28  0.00143 **
## Year2003         1.46508    0.25488    5.75  8.9e-08 ***
## Year2004         0.98218    0.25325    3.88  0.00018 ***
## Year2005         0.73684    0.33281    2.21  0.02899 *
## Year2006         1.09383    0.36722    2.98  0.00360 **
## Year2007         0.34281    0.50203    0.68  0.49621
## Year2008         0.25653    0.27126    0.95  0.34649
## Year2009         0.00286    0.20389    0.01  0.98882
## Year2010         0.18851    0.23077    0.82  0.41586
## Year2011        -0.01940    0.20349   -0.10  0.92421
## Year2012         0.10567    0.22059    0.48  0.63290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.474, Adjusted R-squared:  0.389
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(91,115) are outliers with |weight| = 0 ( < 0.00081);
## 6 weights are ~ = 1. The remaining 115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0407 0.9040 0.9480 0.8820 0.9800 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      8.13e-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	1.875	1	1.369
##	Year	1.875	15	1.021



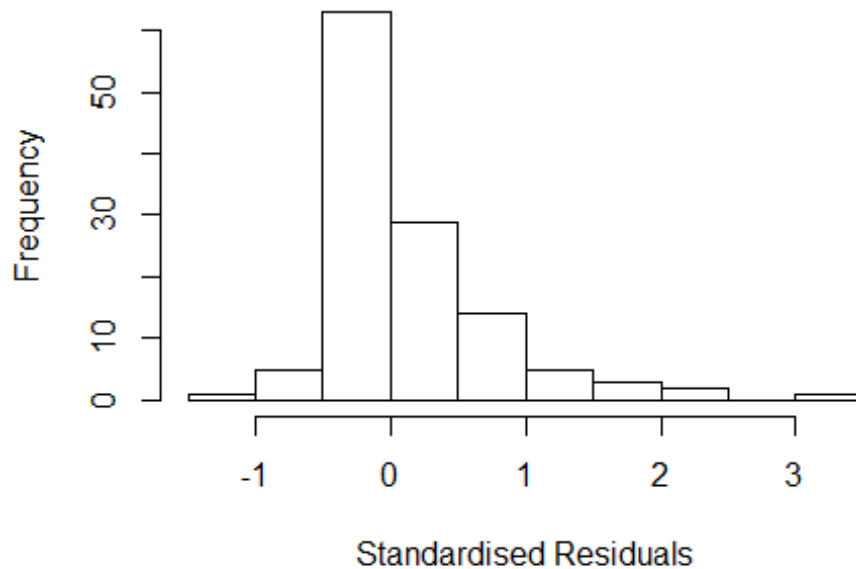
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 174 84860317886 3.474 2012      2914      1      3.128
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.217 -0.324 -0.182  0.460  3.150
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.19634    0.17529   1.12  0.26519
## FirstAuthorFemale1 0.00474    0.13495   0.04  0.97202
## Year1997        1.55741    0.48264   3.23  0.00167 **
## Year1998        0.65091    0.21160   3.08  0.00267 **
## Year2000        0.79259    0.37094   2.14  0.03492 *
## Year2001       -0.20109    0.21160  -0.95  0.34411
## Year2002        1.26708    0.40017   3.17  0.00202 **
## Year2003        1.47139    0.27907   5.27  7.2e-07 ***
```

```

## Year2004          1.02106      0.27073      3.77  0.00027 ***
## Year2005          0.74004      0.32255      2.29  0.02374 *
## Year2006          1.01611      0.35989      2.82  0.00568 **
## Year2007          0.34105      0.45928      0.74  0.45938
## Year2008          0.24167      0.31668      0.76  0.44708
## Year2009          0.01120      0.22473      0.05  0.96036
## Year2010          0.22180      0.25041      0.89  0.37775
## Year2011          0.00187      0.21642      0.01  0.99314
## Year2012          0.12300      0.23080      0.53  0.59520
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.472, Adjusted R-squared:  0.393
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(91,115) are outliers with |weight| = 0 ( < 0.00081);
## 6 weights are ~= 1. The remaining 115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.038  0.881   0.947   0.876   0.981   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          8.13e-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 2.01 1           1.418
## Year             2.01 15           1.024

```

Residuals from last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 174 84860317886 3.474 2012      2914      1      3.128
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.298 -0.337 -0.118  0.451  3.125
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1660    0.1473   1.13  0.26228
## LastAuthorFemale1 0.1052    0.1241   0.85  0.39849
## Year1997        1.4873    0.4680   3.18  0.00195 **
## Year1998        0.5808    0.1739   3.34  0.00115 **
## Year2000        0.7184    0.3774   1.90  0.05969 .
## Year2001       -0.2712    0.1739  -1.56  0.12177
## Year2002        1.1986    0.3784   3.17  0.00201 **
## Year2003        1.4258    0.2541   5.61  1.6e-07 ***
## Year2004        0.9510    0.2421   3.93  0.00015 ***
## Year2005        0.7038    0.3054   2.30  0.02313 *
## Year2006        1.0266    0.3092   3.32  0.00123 **
## Year2007        0.2848    0.4988   0.57  0.56915
```



```

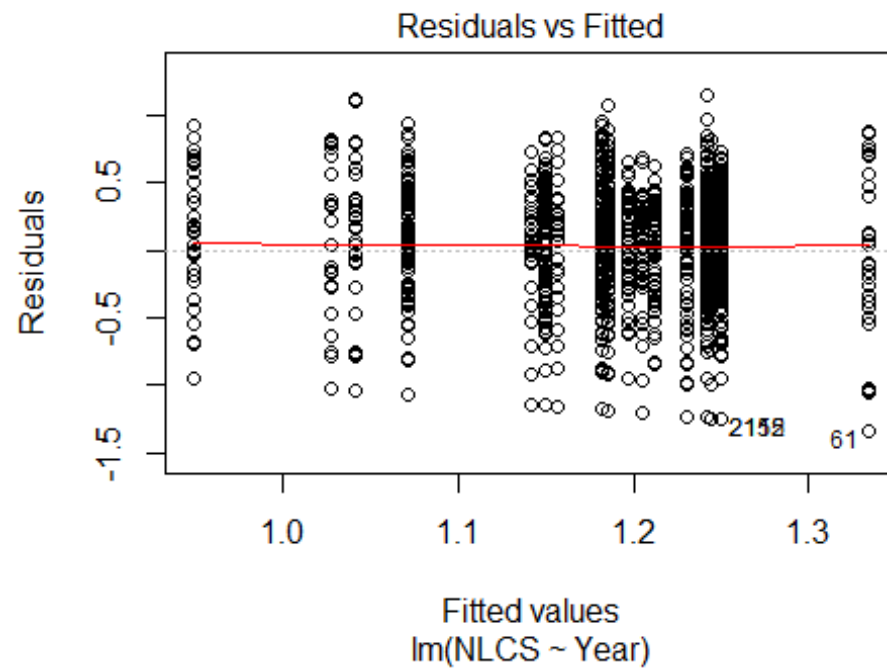
## Year2008          0.2070      0.2712      0.76  0.44701
## Year2009          -0.0202      0.1966     -0.10  0.91827
## Year2010           0.1592      0.2298      0.69  0.48990
## Year2011          -0.0476      0.1937     -0.25  0.80628
## Year2012           0.0783      0.2127      0.37  0.71348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.484, Adjusted R-squared:  0.406
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(91,115) are outliers with |weight| = 0 ( < 0.00081);
## 5 weights are ~= 1. The remaining 116 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0327 0.8770 0.9470 0.8730 0.9780 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      8.13e-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: 123"
## [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]"
##
## 2011
## 1
##
## 2011
## 1
##
## 2011
## 1

```

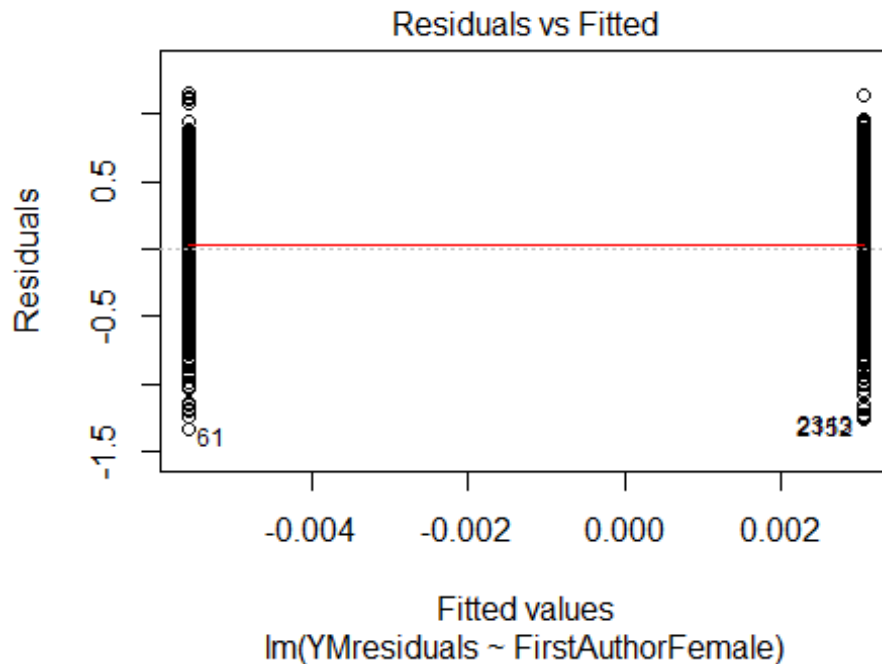
```

## [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: 1"
## [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: 1"
## [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
##    64   49   59   51   73  103   96   72   72   66  102  146  182  199  237
## 2011 2012
##   256  267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    36   28   39   22   37   30   54   51   48   48   77  106  132  138  169
## 2011 2012
##   193  184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    34   27   38   19   37   29   51   43   37   35   68   95  120  125  159
## 2011 2012
##   173  174
## [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

```

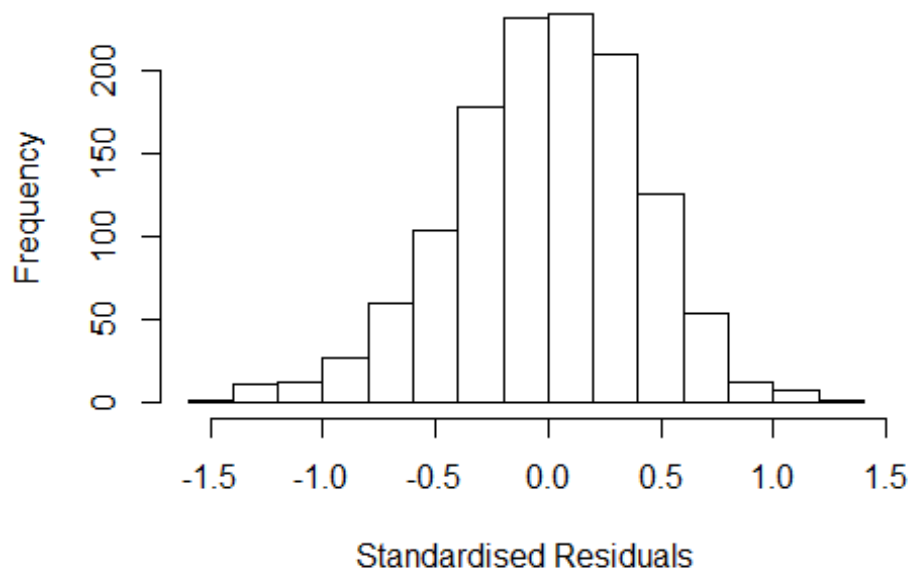


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.70927259188092"
## [1] "Male first author team size 2018 geometric mean: 4.41522518177402"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.59364168813942"
## [1] "Male last author team size 2018 geometric mean: 4.77871593207125"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8000, 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 1.204 1      1.097
## LastAuthorFemale  1.164 1      1.079
## UniqueAuthors    1.434 4      1.046
## Year              1.591 16     1.015
```

Residuals from first and last author and team size



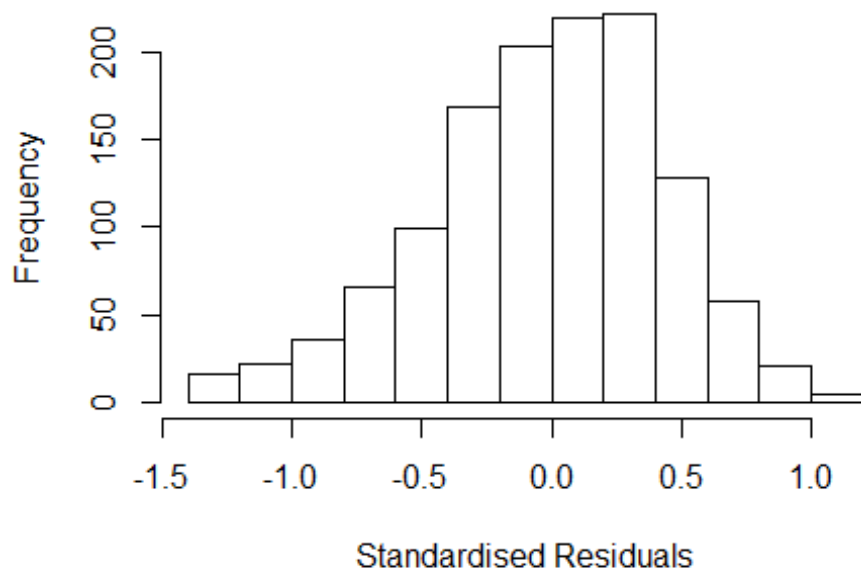
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41500 -0.27908  0.00807  0.27082  1.31588
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0090     0.1184   8.52 < 2e-16 ***
## FirstAuthorFemale1 -0.0240     0.0275  -0.87  0.3835
## LastAuthorFemale1 -0.0653     0.0260  -2.51  0.0122 *
## UniqueAuthors2     0.3593     0.0636   5.65 2.0e-08 ***
## UniqueAuthors3     0.5178     0.0617   8.40 < 2e-16 ***
## UniqueAuthors4     0.4776     0.0619   7.71 2.5e-14 ***
## UniqueAuthors5     0.5850     0.0587   9.96 < 2e-16 ***
## Year1997          -0.2276     0.1454  -1.57  0.1176
## Year1998          -0.2663     0.1531  -1.74  0.0821 .
## Year1999          -0.2508     0.1664  -1.51  0.1320
```

```

## Year2000          -0.2973      0.1433    -2.08    0.0382 *
## Year2001          -0.1570      0.1356    -1.16    0.2473
## Year2002          -0.1325      0.1220    -1.09    0.2776
## Year2003          -0.2161      0.1178    -1.83    0.0669 .
## Year2004          -0.2740      0.1195    -2.29    0.0220 *
## Year2005          -0.1790      0.1226    -1.46    0.1446
## Year2006          -0.1403      0.1174    -1.19    0.2324
## Year2007          -0.3218      0.1164    -2.76    0.0058 **
## Year2008          -0.2745      0.1140    -2.41    0.0162 *
## Year2009          -0.2196      0.1130    -1.94    0.0522 .
## Year2010          -0.2594      0.1116    -2.32    0.0203 *
## Year2011          -0.1864      0.1111    -1.68    0.0938 .
## Year2012          -0.2094      0.1119    -1.87    0.0614 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.164, Adjusted R-squared:  0.15
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1165 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.206  0.870  0.950  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      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 1.183 1      1.088
## LastAuthorFemale  1.139 1      1.067
## Year              1.213 16      1.006

```

Residuals from first and last author



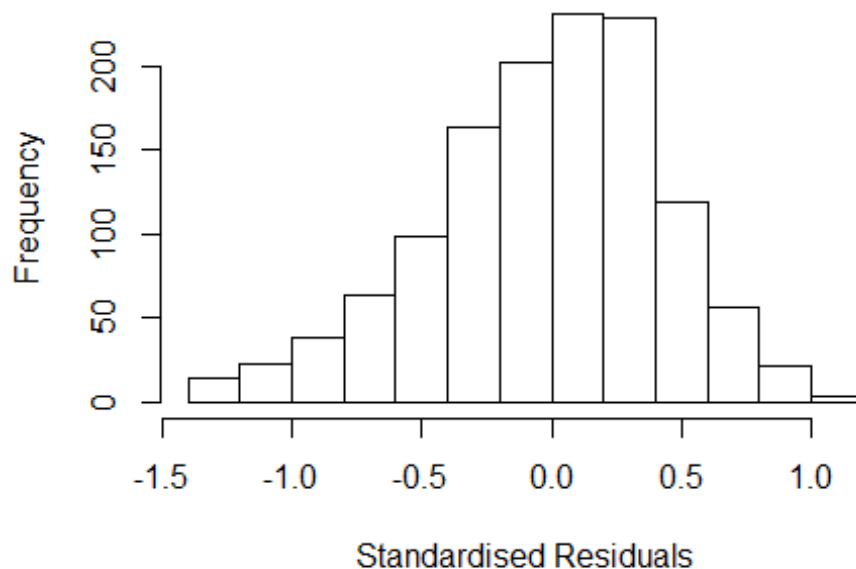
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3933 -0.2986 0.0183 0.2923 1.1034
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3933 0.1166 11.94 <2e-16 ***
## FirstAuthorFemale1 0.0225 0.0287 0.78 0.4338
## LastAuthorFemale1 -0.0869 0.0279 -3.12 0.0019 **
## Year1997 -0.1707 0.1556 -1.10 0.2730
## Year1998 -0.2306 0.1741 -1.32 0.1856
## Year1999 -0.3954 0.2026 -1.95 0.0512 .
## Year2000 -0.4067 0.1579 -2.58 0.0101 *
## Year2001 -0.1805 0.1453 -1.24 0.2147
## Year2002 -0.0849 0.1318 -0.64 0.5198
## Year2003 -0.1602 0.1263 -1.27 0.2048
## Year2004 -0.1831 0.1287 -1.42 0.1551
## Year2005 -0.1415 0.1329 -1.06 0.2873
```

```

## Year2006          -0.0708      0.1280   -0.55   0.5803
## Year2007          -0.2927      0.1288   -2.27   0.0232 *
## Year2008          -0.2038      0.1233   -1.65   0.0985 .
## Year2009          -0.1625      0.1233   -1.32   0.1878
## Year2010          -0.1774      0.1208   -1.47   0.1422
## Year2011          -0.1013      0.1197   -0.85   0.3975
## Year2012          -0.1116      0.1208   -0.92   0.3556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.0376, Adjusted R-squared:  0.0237
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1150 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.266  0.866  0.950  0.898  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      7.91e-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 1.14 1      1.068
## Year              1.14 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
## -1.3801 -0.2924 0.0219 0.2881 1.1381
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.38007 0.11546 11.95 <2e-16 ***
## FirstAuthorFemale1 0.00442 0.02806 0.16 0.875
## Year1997 -0.17376 0.15678 -1.11 0.268
## Year1998 -0.22310 0.17346 -1.29 0.199
## Year1999 -0.38715 0.20300 -1.91 0.057 .
## Year2000 -0.41314 0.15543 -2.66 0.008 **
## Year2001 -0.18097 0.14380 -1.26 0.208
## Year2002 -0.08811 0.13048 -0.68 0.500
## Year2003 -0.15088 0.12535 -1.20 0.229
## Year2004 -0.17825 0.12871 -1.38 0.166
## Year2005 -0.14896 0.13191 -1.13 0.259
## Year2006 -0.08621 0.12630 -0.68 0.495
```

```

## Year2007          -0.30125    0.12842   -2.35    0.019 *
## Year2008          -0.21532    0.12245   -1.76    0.079 .
## Year2009          -0.16966    0.12247   -1.39    0.166
## Year2010          -0.19314    0.11982   -1.61    0.107
## Year2011          -0.11869    0.11866   -1.00    0.317
## Year2012          -0.13312    0.11947   -1.11    0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.0162
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1154 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.276  0.864  0.948  0.898  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      7.91e-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 1.088 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
## -1.4015 -0.2960  0.0188  0.2962  1.0880

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4015     0.1156   12.12 <2e-16 ***
## LastAuthorFemale1 -0.0829     0.0272   -3.05  0.0024 **
## Year1997          -0.1725     0.1547   -1.12  0.2649
## Year1998          -0.2390     0.1714   -1.39  0.1635
## Year1999          -0.4011     0.2016   -1.99  0.0468 *
## Year2000          -0.4082     0.1574   -2.59  0.0096 **
## Year2001          -0.1794     0.1444   -1.24  0.2143
## Year2002          -0.0827     0.1308   -0.63  0.5273
## Year2003          -0.1552     0.1259   -1.23  0.2179
## Year2004          -0.1768     0.1281   -1.38  0.1678
## Year2005          -0.1400     0.1322   -1.06  0.2896
## Year2006          -0.0704     0.1271   -0.55  0.5796
## Year2007          -0.2892     0.1283   -2.25  0.0244 *
## Year2008          -0.1998     0.1228   -1.63  0.1041
## Year2009          -0.1567     0.1228   -1.28  0.2023
## Year2010          -0.1710     0.1204   -1.42  0.1557
## Year2011          -0.0945     0.1191   -0.79  0.4278
## Year2012          -0.1045     0.1203   -0.87  0.3853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.432
## Multiple R-squared:  0.0372, Adjusted R-squared:  0.024
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1151 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.271  0.871  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      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 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    2    5    3    2    5    5    4    2    4    4    7    9    9    10
## 2011 2012
##   14   17
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    0    3    2    2    4    5    4    2    3    3    3    9    9    10
## 2011 2012
##   10   15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    0    2    2    2    4    5    4    2    3    3    3    9    9    8
## 2011 2012
##    9   14
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.44718220016196"
## [1] "Male first author team size 2018 geometric mean: 5"

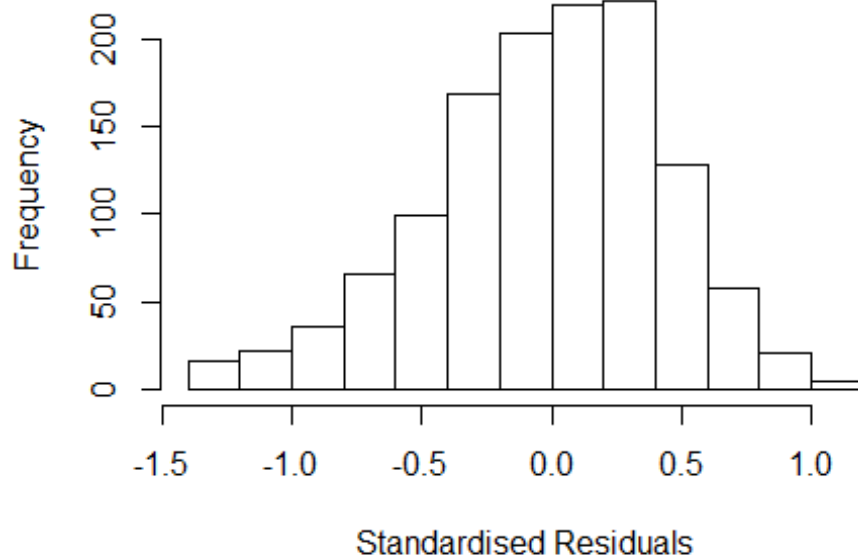
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.66663527497234"
## [1] "Male last author team size 2018 geometric mean: 6.11690884906252"

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

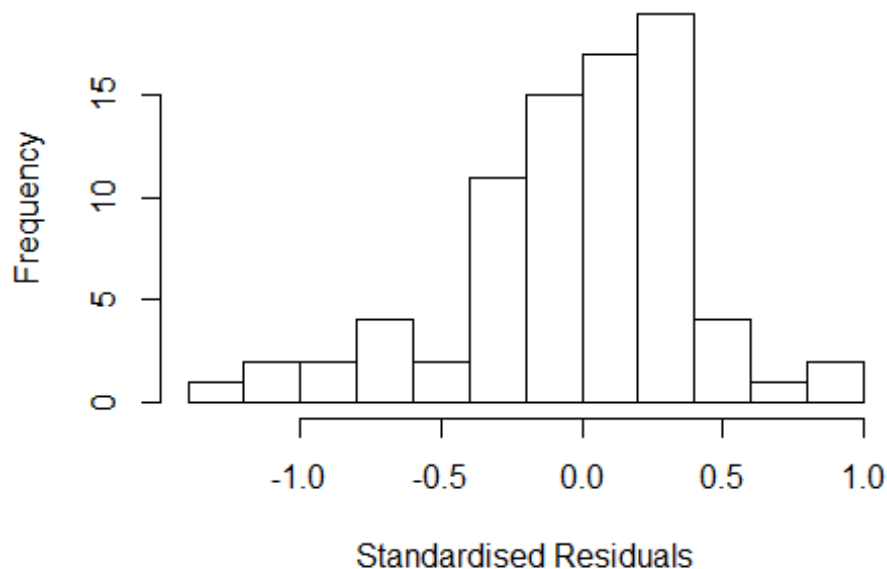
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, 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    1.684  1      1.298
## LastAuthorFemale     3.473  1      1.864
## UniqueAuthors    1388.553  4      2.471
## Year           2339.849 15      1.295
```

Residuals from first and last author and team size



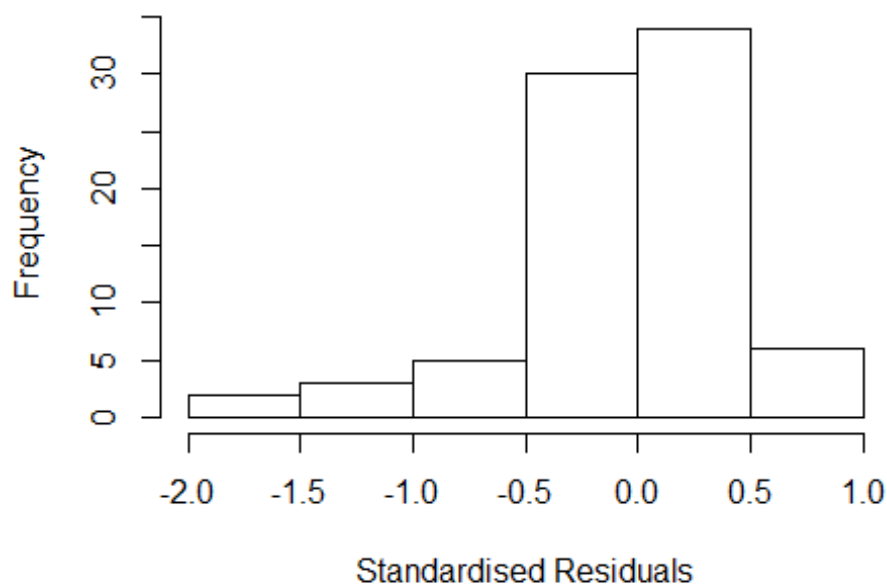
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2632 -0.2289 0.0479 0.2311 0.8405
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93171 0.27665 3.37 0.0014 **
## FirstAuthorFemale1 0.46085 0.20460 2.25 0.0281 *
## LastAuthorFemale1 0.20023 0.16282 1.23 0.2237
## UniqueAuthors2 0.02520 0.22751 0.11 0.9122
## UniqueAuthors3 0.65233 0.24612 2.65 0.0103 *
## UniqueAuthors4 0.16686 0.22982 0.73 0.4707
## UniqueAuthors5 0.51797 0.24255 2.14 0.0370 *
## Year1998 -0.43680 0.26021 -1.68 0.0986 .
## Year1999 -0.18630 0.26510 -0.70 0.4850
## Year2000 -0.67773 0.21717 -3.12 0.0028 **
```

```

## Year2001      -0.20064    0.17446   -1.15    0.2548
## Year2002      -0.55578    0.33687   -1.65    0.1044
## Year2003      -0.76825    0.24921   -3.08    0.0031 **
## Year2004      -0.90045    0.32131   -2.80    0.0069 **
## Year2005      -0.48344    0.28941   -1.67    0.1002
## Year2006      -0.45739    0.24350   -1.88    0.0654 .
## Year2007      -0.00775    0.14478   -0.05    0.9575
## Year2008      -0.31888    0.20977   -1.52    0.1339
## Year2009      -0.15456    0.32399   -0.48    0.6351
## Year2010      -0.52825    0.25618   -2.06    0.0437 *
## Year2011      -0.66144    0.42781   -1.55    0.1275
## Year2012      -0.54145    0.24778   -2.19    0.0329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.376, Adjusted R-squared:  0.15
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 74 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.343  0.921  0.968  0.910  0.988  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      1.25e-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 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.491 1 1.221
## LastAuthorFemale 2.702 1 1.644
## Year 4.024 15 1.048

```

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
## -1.76550 -0.21336 0.00625 0.20267 0.89006
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09588 0.25849 4.24 7.6e-05 ***
## FirstAuthorFemale1 0.42423 0.21307 1.99 0.05089 .
## LastAuthorFemale1 0.09789 0.14805 0.66 0.51092
## Year1998 -0.46200 0.12631 -3.66 0.00053 ***
## Year1999 -0.21150 0.13611 -1.55 0.12529
## Year2000 -0.61950 0.00884 -70.08 < 2e-16 ***
## Year2001 -0.17312 0.11894 -1.46 0.15057
## Year2002 -0.25996 0.26471 -0.98 0.32988
## Year2003 -0.35893 0.11538 -3.11 0.00282 **
## Year2004 -0.34050 0.22764 -1.50 0.13979
## Year2005 -0.14535 0.13478 -1.08 0.28502
## Year2006 -0.09261 0.31000 -0.30 0.76613
```

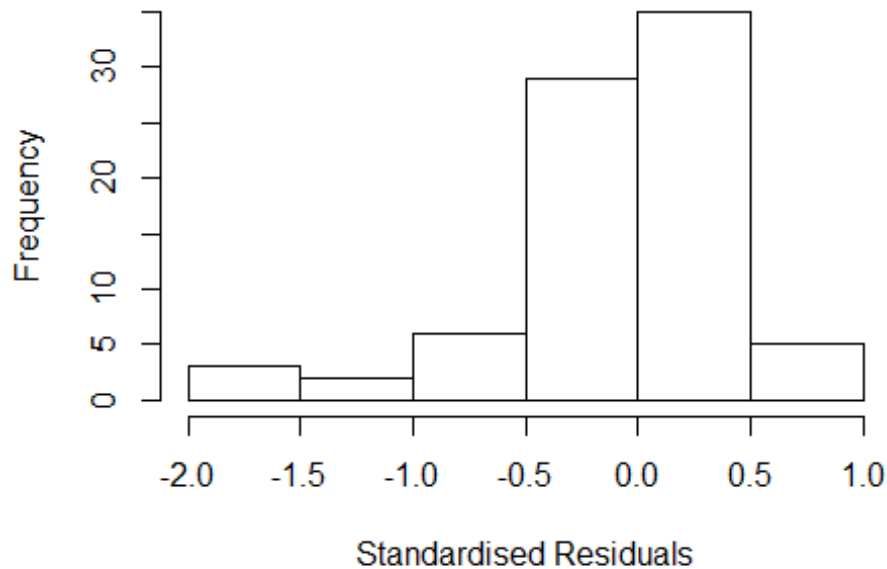


```

## Year2007          -0.05198      0.12490      -0.42      0.67873
## Year2008          -0.01088      0.11077      -0.10      0.92208
## Year2009           0.24539      0.16764       1.46      0.14831
## Year2010          -0.29121      0.21307      -1.37      0.17664
## Year2011          -0.04244      0.28844      -0.15      0.88352
## Year2012          -0.22106      0.13395      -1.65      0.10394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.233, Adjusted R-squared:  0.0225
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 70 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0345 0.8920 0.9720 0.8770 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      1.25e-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 1.498 1          1.224
## Year              1.498 15          1.014

```

Residuals from first author



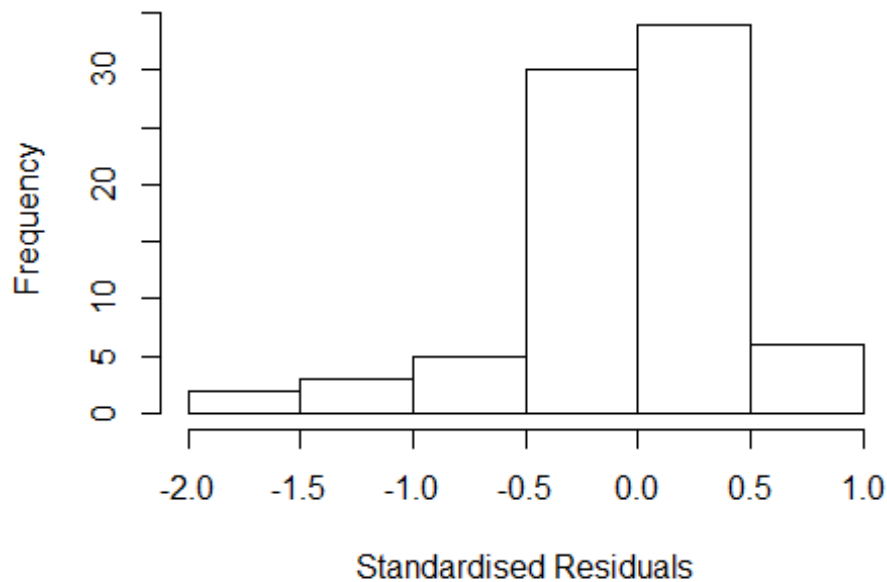
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.85187 -0.21482 0.00625 0.18227 0.90831
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18834 0.21104 5.63 4.5e-07 ***
## FirstAuthorFemale1 0.42966 0.21104 2.04 0.04597 *
## Year1998 -0.46200 0.12655 -3.65 0.00053 ***
## Year1999 -0.21150 0.13641 -1.55 0.12603
## Year2000 -0.61950 0.00884 -70.08 < 2e-16 ***
## Year2001 -0.17336 0.11922 -1.45 0.15090
## Year2002 -0.27999 0.25361 -1.10 0.27379
## Year2003 -0.38242 0.12651 -3.02 0.00362 **
## Year2004 -0.34050 0.22901 -1.49 0.14204
## Year2005 -0.20890 0.10691 -1.95 0.05514 .
## Year2006 -0.12649 0.28307 -0.45 0.65653
## Year2007 -0.08642 0.11422 -0.76 0.45212
```

```

## Year2008          -0.02613      0.11247      -0.23      0.81702
## Year2009          0.23387      0.14412       1.62      0.10964
## Year2010         -0.32731      0.19308      -1.70      0.09498 .
## Year2011         -0.04009      0.29146      -0.14      0.89102
## Year2012         -0.23819      0.13331      -1.79      0.07880 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.242, Adjusted R-squared:  0.0498
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 71 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0027 0.8880 0.9740 0.8720 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      1.25e-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 2.781 1          1.668
## Year            2.781 15          1.035

```

Residuals from last author



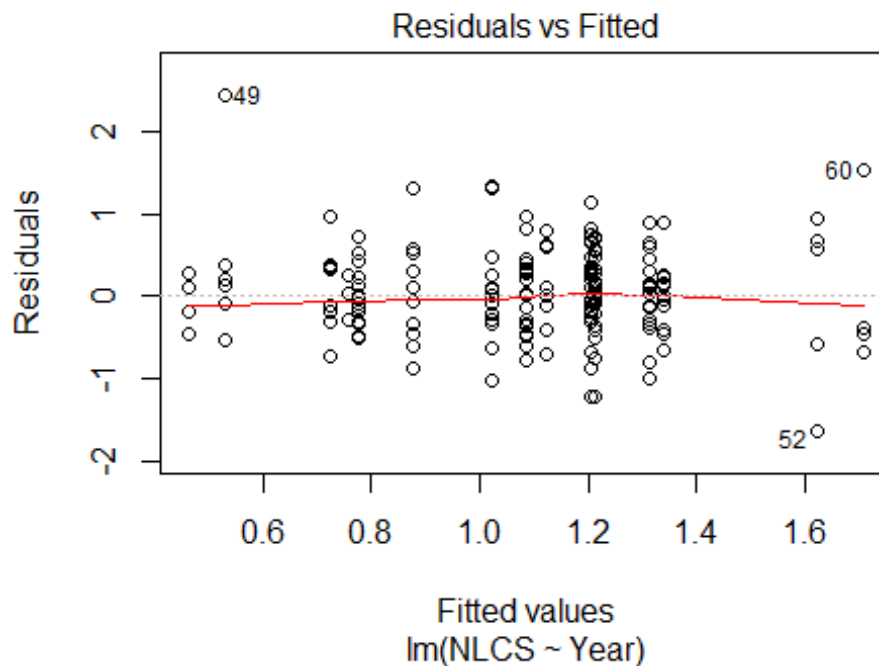
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5943 -0.2254 0.0109 0.2501 0.9319
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.49858 0.16352 9.16 3.4e-13 ***
## LastAuthorFemale1 0.11942 0.16352 0.73 0.4679
## Year1998 -0.46200 0.12581 -3.67 0.0005 ***
## Year1999 -0.21150 0.13548 -1.56 0.1235
## Year2000 -0.61950 0.00884 -70.08 < 2e-16 ***
## Year2001 -0.17262 0.11834 -1.46 0.1496
## Year2002 -0.28232 0.27419 -1.03 0.3071
## Year2003 -0.35321 0.11359 -3.11 0.0028 **
## Year2004 -0.34050 0.22484 -1.51 0.1349
## Year2005 -0.13127 0.14262 -0.92 0.3609
## Year2006 -0.09035 0.30921 -0.29 0.7711
## Year2007 -0.04439 0.12632 -0.35 0.7265
```

```

## Year2008          -0.08735      0.10533      -0.83      0.4101
## Year2009           0.09571      0.16578       0.58      0.5658
## Year2010          -0.35090      0.21822     -1.61      0.1128
## Year2011          -0.23080      0.25831     -0.89      0.3750
## Year2012          -0.26060      0.19129     -1.36      0.1780
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.12,   Adjusted R-squared:  -0.103
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 71 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.174  0.908  0.961  0.886  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      1.25e-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: 80"
## [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
##    5    9    6   10    2   13    6    4   12   12   10   17   19   24   23
## 2011 2012
##   29   30
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    8    5    7    0   12    5    4   11   11    9   12   17   20   19
## 2011 2012
##   26   28

```

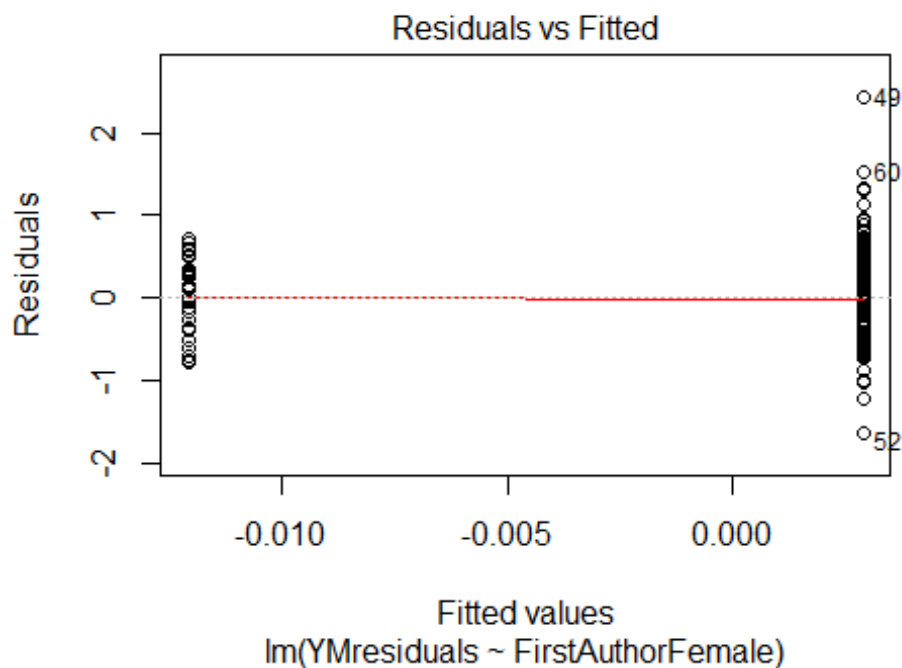
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    7    5    6    0   12    5    4   11   10    9   12   16   19   19
## 2011 2012
##   24   24
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 24, df = 15, p-value = 0.07
```



```
##
## 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: 3.20461490314515"
## [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 = 40, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.27381210096172"
## [1] "Male last author team size 2018 geometric mean: 3"

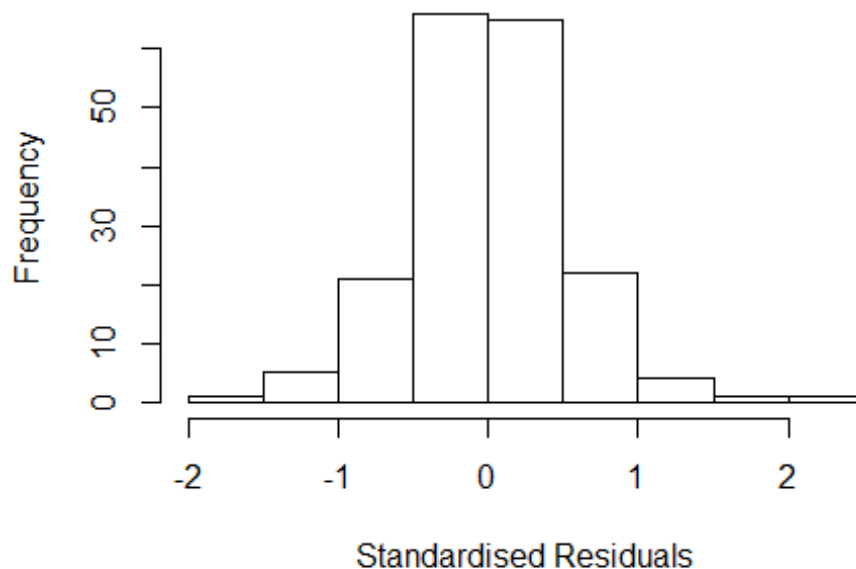
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 74, 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"
##
```

	GVI	F	Df	GVI ^{1/(2*Df)}
FirstAuthorFemale	1.846	1		1.359
LastAuthorFemale	2.076	1		1.441
UniqueAuthors	13.599	4		1.386
Year	22.557	15		1.109

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
## -1.616472 -0.282752 0.000915 0.321656 2.476484
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6344 0.2596 2.44 0.0156 *
## FirstAuthorFemale1 -0.0221 0.1052 -0.21 0.8338
## LastAuthorFemale1 -0.0444 0.0981 -0.45 0.6514
## UniqueAuthors2 0.2839 0.0997 2.85 0.0050 **
## UniqueAuthors3 0.2508 0.1450 1.73 0.0855 .
## UniqueAuthors4 0.2126 0.1799 1.18 0.2391
## UniqueAuthors5 0.5006 0.2127 2.35 0.0198 *
## Year1997 0.0645 0.2611 0.25 0.8051
## Year1998 -0.2922 0.2334 -1.25 0.2123
## Year1999 -0.0613 0.2322 -0.26 0.7920
```

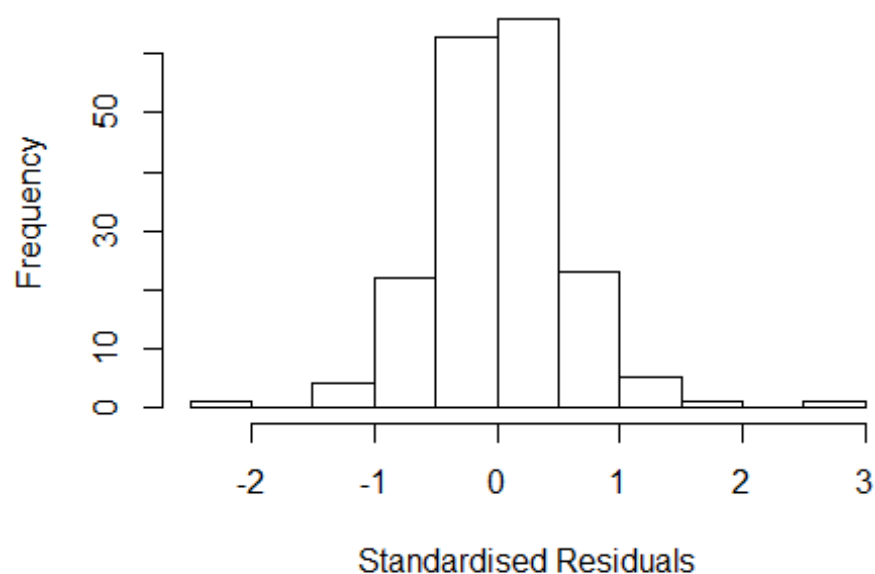


```

## Year2001          -0.3372      0.2495    -1.35    0.1783
## Year2002           1.0485      0.6915     1.52    0.1314
## Year2003           0.6489      0.4051     1.60    0.1111
## Year2004           0.0171      0.2752     0.06    0.9506
## Year2005           0.6271      0.2328     2.69    0.0078 **
## Year2006          -0.1060      0.2753    -0.38    0.7008
## Year2007           0.3722      0.2797     1.33    0.1852
## Year2008           0.2372      0.2952     0.80    0.4227
## Year2009           0.2753      0.2428     1.13    0.2584
## Year2010           0.5622      0.2461     2.28    0.0236 *
## Year2011           0.4742      0.2452     1.93    0.0548 .
## Year2012           0.2975      0.2482     1.20    0.2324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.358, Adjusted R-squared:  0.276
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## observation 32 is an outlier with |weight| = 0 ( < 0.00054);
## 15 weights are ~= 1. The remaining 170 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.173  0.883  0.960  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          5.38e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
## trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.653 1          1.286
## LastAuthorFemale  1.701 1          1.304
## Year              1.810 15          1.020

```

Residuals from first and last author



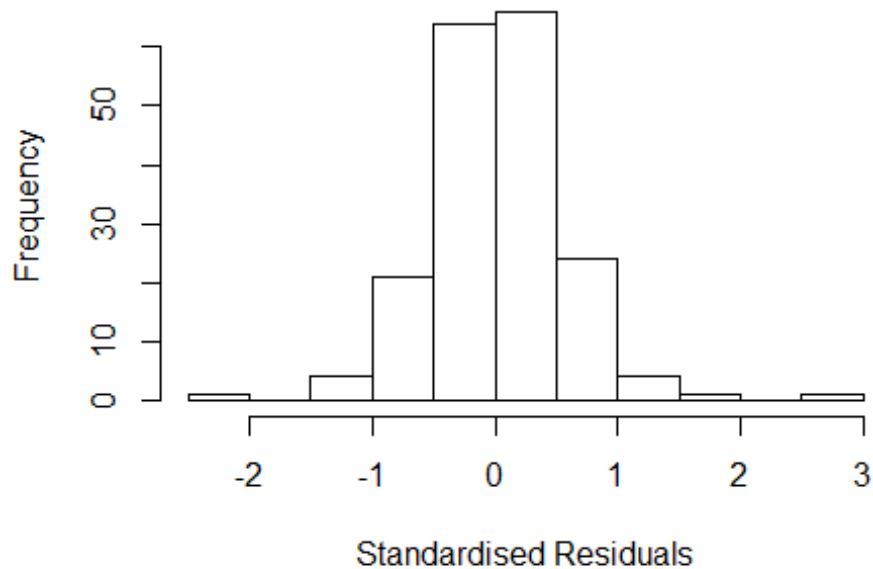
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 49 0035453382 2.958 2001      2906      3      2.686
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0563 -0.2997  0.0197  0.2985  2.6862
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.84401    0.15540   5.43 1.9e-07 ***
## FirstAuthorFemale1  0.00738    0.10274   0.07  0.94283
## LastAuthorFemale1 -0.09198    0.09323  -0.99  0.32525
## Year1997         0.00879    0.18856   0.05  0.96286
## Year1998        -0.30768    0.17834  -1.73  0.08633 .
## Year1999        -0.07009    0.17405  -0.40  0.68768
## Year2001        -0.48761    0.17081  -2.85  0.00485 **
## Year2002         1.29690    0.43897   2.95  0.00358 **
## Year2003         0.46514    0.19347   2.40  0.01730 *
## Year2004         0.03549    0.22761   0.16  0.87629
## Year2005         0.59114    0.17528   3.37  0.00092 ***
## Year2006        -0.07814    0.23083  -0.34  0.73540
```

```

## Year2007          0.34058      0.22299      1.53  0.12856
## Year2008          0.18974      0.19928      0.95  0.34240
## Year2009          0.26187      0.17675      1.48  0.14033
## Year2010          0.55062      0.17044      3.23  0.00149 **
## Year2011          0.50459      0.17865      2.82  0.00531 **
## Year2012          0.46456      0.16244      2.86  0.00478 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.324, Adjusted R-squared:  0.255
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 32 is an outlier with |weight| = 0 ( < 0.00054);
## 12 weights are ~= 1. The remaining 173 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0554 0.8960 0.9610 0.9040 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      5.38e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.369 1      1.170
## Year              1.369 15      1.011

```

Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 49 0035453382 2.958 2001      2906      3      2.686
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0701 -0.2935  0.0155  0.3060  2.6645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7770     0.1527   5.09 9.5e-07 ***
## FirstAuthorFemale1 -0.0294     0.0933  -0.32  0.75277
## Year1997          0.0353     0.1918   0.18  0.85403
## Year1998         -0.2924     0.1872  -1.56  0.12020
## Year1999         -0.0535     0.1899  -0.28  0.77850
## Year2001         -0.4541     0.1781  -2.55  0.01168 *
## Year2002          1.3225     0.4555   2.90  0.00418 **
## Year2003          0.5043     0.2096   2.41  0.01719 *
## Year2004          0.0635     0.2384   0.27  0.79027
## Year2005          0.6361     0.1755   3.62  0.00038 ***
## Year2006         -0.0452     0.2358  -0.19  0.84817
## Year2007          0.3698     0.2262   1.64  0.10386
```

```

## Year2008          0.2248      0.2007      1.12  0.26424
## Year2009          0.2946      0.1794      1.64  0.10251
## Year2010          0.5687      0.1755      3.24  0.00144 **
## Year2011          0.5359      0.1831      2.93  0.00389 **
## Year2012          0.4983      0.1630      3.06  0.00260 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.32,   Adjusted R-squared:  0.256
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 32 is an outlier with |weight| = 0 ( < 0.00054);
## 12 weights are ~= 1. The remaining 173 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.053  0.875   0.964   0.904   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      5.38e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.387  1          1.178
## Year              1.387 15          1.011

## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 49 0035453382 2.958 2001      2906      3      2.686
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -2.0617 -0.2975  0.0234  0.2947  2.6876

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.84685    0.15323    5.53 1.2e-07 ***
## LastAuthorFemale1 -0.08981    0.08448   -1.06 0.28924
## Year1997         0.00869    0.18898    0.05 0.96339
## Year1998        -0.30781    0.17902   -1.72 0.08736 .
## Year1999        -0.06885    0.17654   -0.39 0.69703
## Year2001        -0.48661    0.17235   -2.82 0.00532 **
## Year2002         1.30468    0.43121    3.03 0.00287 **
## Year2003         0.45934    0.18774    2.45 0.01544 *
## Year2004         0.03506    0.22869    0.15 0.87832
## Year2005         0.59251    0.17598    3.37 0.00094 ***
## Year2006        -0.07751    0.23209   -0.33 0.73883
## Year2007         0.34274    0.22407    1.53 0.12798
## Year2008         0.19083    0.19842    0.96 0.33754
## Year2009         0.26269    0.17780    1.48 0.14142
## Year2010         0.55249    0.16767    3.30 0.00120 **
## Year2011         0.50781    0.17827    2.85 0.00494 **
## Year2012         0.46632    0.16101    2.90 0.00428 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.326, Adjusted R-squared:  0.262
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 32 is an outlier with |weight| = 0 ( < 0.00054);
## 12 weights are ~= 1. The remaining 173 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0477 0.8930 0.9610 0.9030 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           5.38e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 186"
## [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]"
##
## 2000 2001 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    1    1    1    6    8    8   12   15    8    6
##
## 2000 2001 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    1    0    5    7    7   11   12    7    5
##
## 2000 2001 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    1    0    4    6    7    9   12    6    4
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.25992104989487"
## [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"
## [1] "Male last author team size 2018 geometric mean: 1.41421356237309"

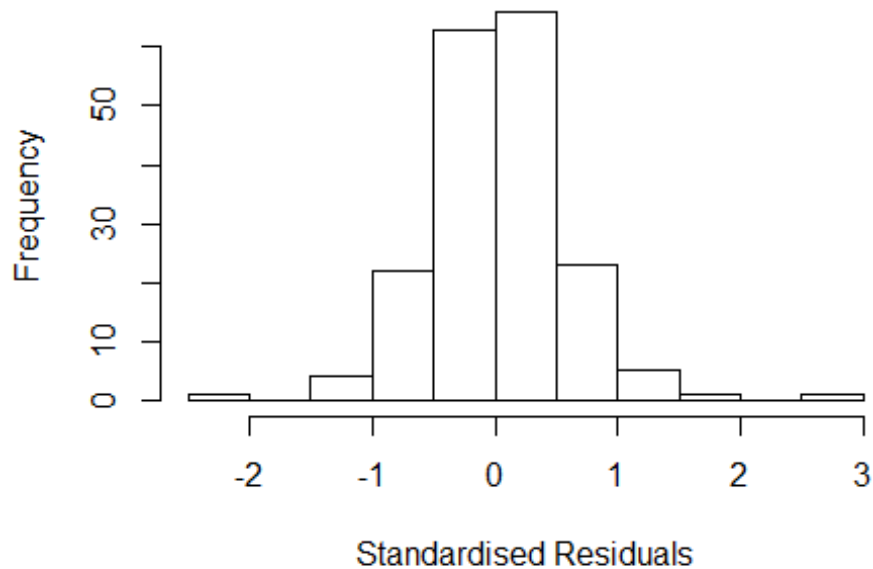
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1, p-value = 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"
## [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

```

Residuals from last author



```
## [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: 49"
## [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
##    3    9    4   14    6   13   29   25   26   26   32   47   54   76   65
## 2011 2012
## 105   77
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    5    2    6    4    3   24   20   21   20   19   34   45   65   55
## 2011 2012
##   86   58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    5    2    5    4    3   20   16   19   18   18   33   41   60   52
## 2011 2012
##   80   52
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
```



```

## [1] "Female first author team size 2018 geometric mean: 3.45882130082181"
## [1] "Male first author team size 2018 geometric mean: 3.79256945098166"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 980, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.5232057854332"
## [1] "Male last author team size 2018 geometric mean: 3.69551710302141"
##
## 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 3.094 1          1.759
## LastAuthorFemale 3.641 1          1.908
## UniqueAuthors    17.739 4          1.433
## Year              27.796 16         1.109

## [1] "List of 0 outliers with 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
## -1.481 -0.305  0.017  0.307  1.642
##
## Coefficients:
##
##          Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.66e+00   8.58e-08  1.94e+07 < 2e-16 ***
## FirstAuthorFemale1 -1.34e-02   5.15e-02 -2.60e-01  0.79440
## LastAuthorFemale1  1.03e-01   5.41e-02  1.90e+00  0.05845 .
## UniqueAuthors2    2.41e-01   8.78e-02  2.74e+00  0.00634 **
## UniqueAuthors3    1.92e-01   8.76e-02  2.20e+00  0.02860 *
## UniqueAuthors4    7.02e-02   1.09e-01  6.40e-01  0.52088
## UniqueAuthors5    9.42e-02   9.51e-02  9.90e-01  0.32267

```

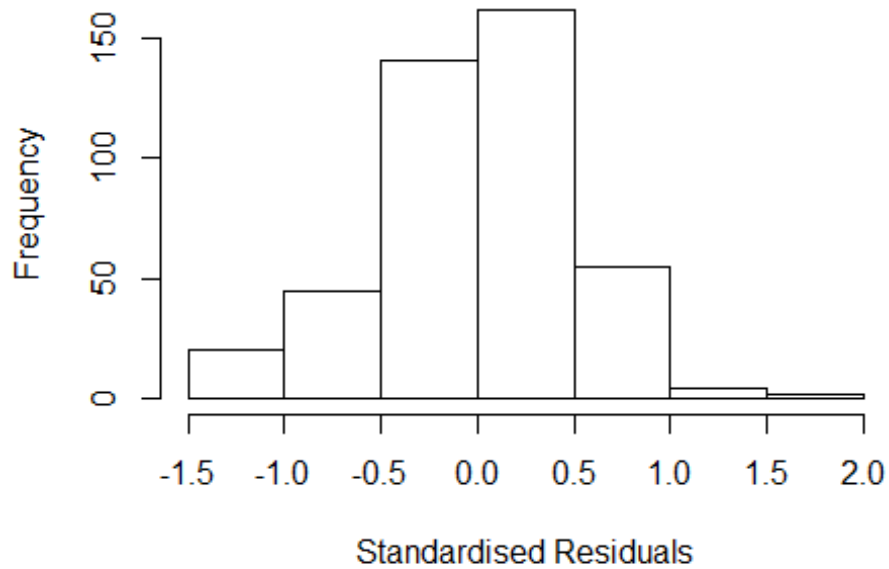
```

## Year1997      -7.08e-01    9.22e-02 -7.67e+00    1.3e-13 ***
## Year1998      -1.63e-01    1.23e-01 -1.32e+00    0.18704
## Year1999      -1.08e+00    2.13e-01 -5.09e+00    5.4e-07 ***
## Year2000      -1.22e+00    3.87e-01 -3.16e+00    0.00169 **
## Year2001      -8.31e-01    5.20e-01 -1.60e+00    0.11071
## Year2002      -6.35e-01    1.51e-01 -4.20e+00    3.3e-05 ***
## Year2003      -3.99e-01    1.07e-01 -3.71e+00    0.00024 ***
## Year2004      -5.83e-01    2.09e-01 -2.79e+00    0.00555 **
## Year2005      -2.68e-01    1.24e-01 -2.16e+00    0.03165 *
## Year2006      -5.67e-01    1.05e-01 -5.41e+00    1.1e-07 ***
## Year2007      -7.46e-01    1.02e-01 -7.31e+00    1.4e-12 ***
## Year2008      -5.36e-01    1.07e-01 -5.00e+00    8.6e-07 ***
## Year2009      -5.37e-01    1.07e-01 -5.03e+00    7.4e-07 ***
## Year2010      -5.85e-01    1.12e-01 -5.21e+00    3.1e-07 ***
## Year2011      -5.25e-01    9.88e-02 -5.32e+00    1.7e-07 ***
## Year2012      -6.93e-01    1.05e-01 -6.61e+00    1.2e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0658
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 392 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.222  0.862  0.956  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           2.33e-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"

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

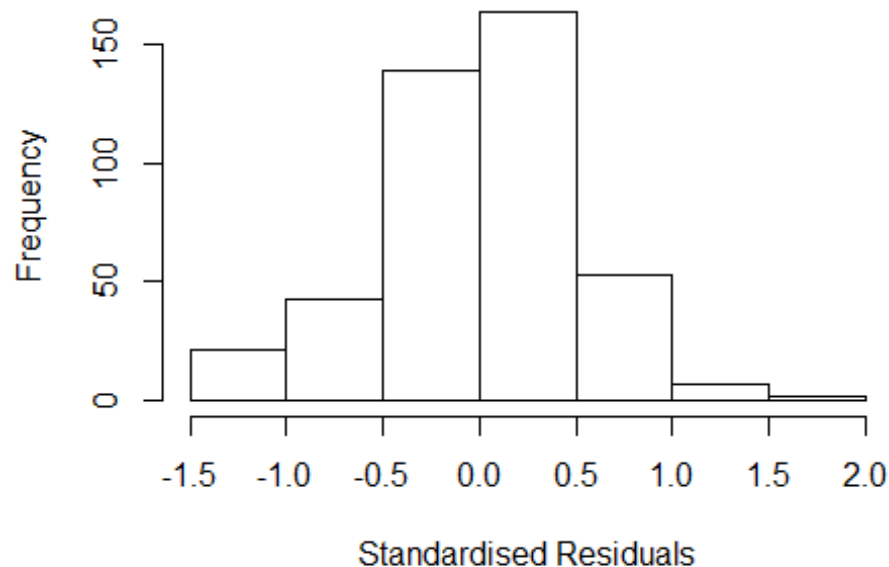
```

Residuals from first and last author and team size



##		GVIF	Df	$GVIF^{1/(2 \cdot Df)}$
##	FirstAuthorFemale	2.846	1	1.687
##	LastAuthorFemale	3.129	1	1.769
##	Year	5.249	16	1.053

Residuals from first and last author



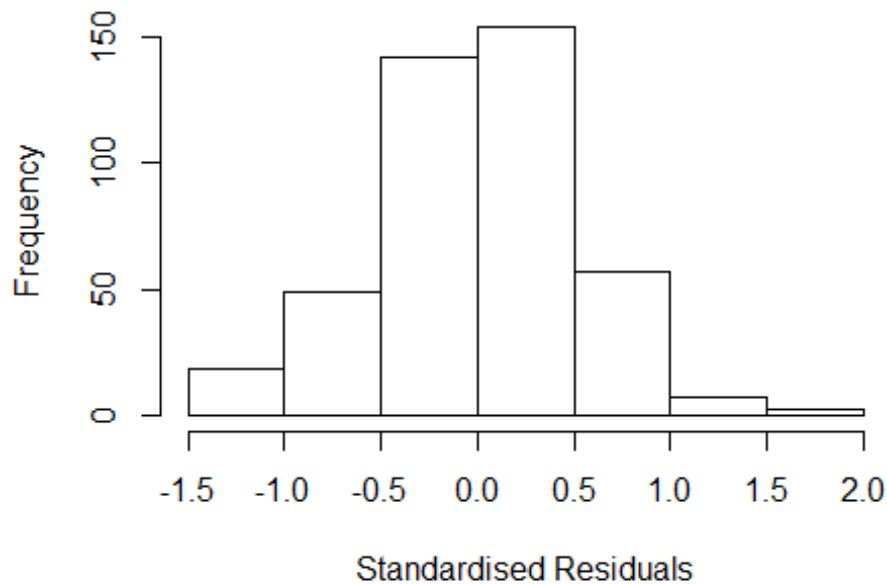
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.406 -0.328 0.032 0.297 1.717
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6630 0.0000 Inf < 2e-16 ***
## FirstAuthorFemale1 -0.0158 0.0523 -0.30 0.7624
## LastAuthorFemale1 0.1248 0.0540 2.31 0.0212 *
## Year1997 -0.5920 0.1094 -5.41 1.1e-07 ***
## Year1998 -0.0520 0.2020 -0.26 0.7971
## Year1999 -0.9141 0.1708 -5.35 1.4e-07 ***
## Year2000 -1.1431 0.3597 -3.18 0.0016 **
## Year2001 -0.7364 0.4188 -1.76 0.0795 .
## Year2002 -0.4834 0.1289 -3.75 0.0002 ***
## Year2003 -0.3040 0.0990 -3.07 0.0023 **
## Year2004 -0.4299 0.1983 -2.17 0.0307 *
## Year2005 -0.1053 0.1045 -1.01 0.3144
```

```

## Year2006          -0.4416      0.0906   -4.88   1.6e-06 ***
## Year2007          -0.6327      0.0897   -7.05   7.5e-12 ***
## Year2008          -0.4172      0.0787   -5.30   1.9e-07 ***
## Year2009          -0.4119      0.0881   -4.68   3.9e-06 ***
## Year2010          -0.4408      0.0895   -4.93   1.2e-06 ***
## Year2011          -0.3819      0.0772   -4.95   1.1e-06 ***
## Year2012          -0.5771      0.0826   -6.99   1.1e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0867, Adjusted R-squared:  0.0466
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 402 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.202  0.879   0.956   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      2.33e-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 2.907 1      1.705
## Year              2.907 16      1.034

```

Residuals from first author



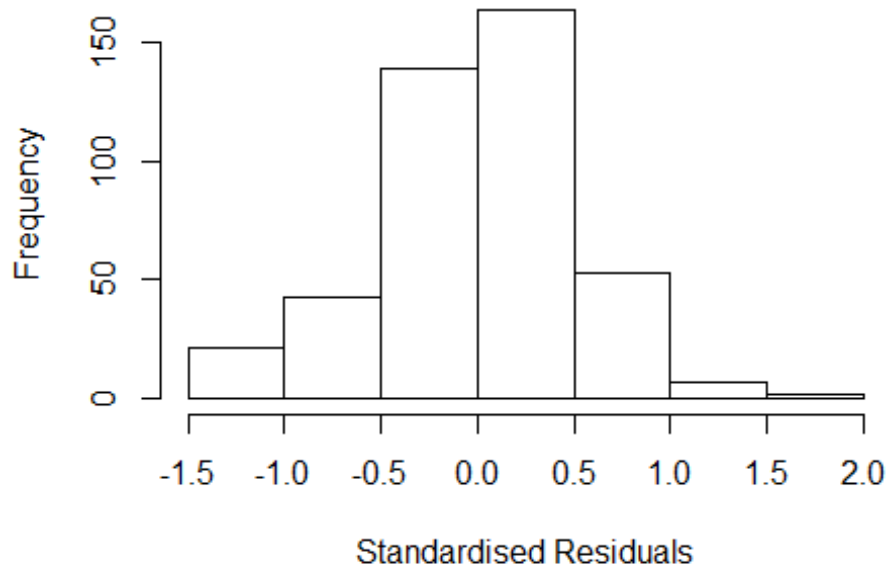
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3220 -0.3493 0.0111 0.3011 1.8081
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.66e+00 2.52e-07 6.59e+06 < 2e-16 ***
## FirstAuthorFemale1 7.14e-03 5.32e-02 1.30e-01 0.8934
## Year1997 -5.48e-01 1.36e-01 -4.03e+00 6.6e-05 ***
## Year1998 -1.07e-03 2.44e-01 0.00e+00 0.9965
## Year1999 -8.96e-01 1.67e-01 -5.37e+00 1.3e-07 ***
## Year2000 -1.09e+00 4.22e-01 -2.59e+00 0.0101 *
## Year2001 -7.08e-01 4.53e-01 -1.56e+00 0.1193
## Year2002 -4.29e-01 1.34e-01 -3.21e+00 0.0014 **
## Year2003 -2.30e-01 9.87e-02 -2.33e+00 0.0202 *
## Year2004 -3.68e-01 1.94e-01 -1.90e+00 0.0581 .
## Year2005 -3.60e-02 9.91e-02 -3.60e-01 0.7168
## Year2006 -3.64e-01 8.28e-02 -4.40e+00 1.4e-05 ***
```

```

## Year2007          -5.43e-01    7.60e-02 -7.15e+00  4.0e-12 ***
## Year2008          -3.69e-01    7.72e-02 -4.78e+00  2.5e-06 ***
## Year2009          -3.57e-01    8.15e-02 -4.38e+00  1.5e-05 ***
## Year2010          -3.77e-01    8.10e-02 -4.65e+00  4.5e-06 ***
## Year2011          -3.48e-01    7.72e-02 -4.51e+00  8.4e-06 ***
## Year2012          -5.19e-01    8.12e-02 -6.39e+00  4.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0741, Adjusted R-squared:  0.0358
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 402 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.158  0.874  0.955  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      2.33e-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 3.077 1      1.754
## Year              3.077 16      1.036

```

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
## -1.3930 -0.3230 0.0253 0.3049 1.7300
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.66e+00 1.56e-07 1.07e+07 < 2e-16 ***
## LastAuthorFemale1 1.22e-01 5.41e-02 2.25e+00 0.02484 *
## Year1997 -5.94e-01 1.08e-01 -5.50e+00 6.7e-08 ***
## Year1998 -5.84e-02 1.95e-01 -3.00e-01 0.76481
## Year1999 -9.21e-01 1.68e-01 -5.49e+00 7.2e-08 ***
## Year2000 -1.15e+00 3.55e-01 -3.23e+00 0.00135 **
## Year2001 -7.51e-01 4.17e-01 -1.80e+00 0.07258 .
## Year2002 -4.90e-01 1.29e-01 -3.81e+00 0.00016 ***
## Year2003 -3.15e-01 9.16e-02 -3.44e+00 0.00065 ***
## Year2004 -4.38e-01 1.99e-01 -2.20e+00 0.02852 *
## Year2005 -1.16e-01 9.69e-02 -1.20e+00 0.23242
## Year2006 -4.47e-01 8.85e-02 -5.05e+00 6.7e-07 ***
```

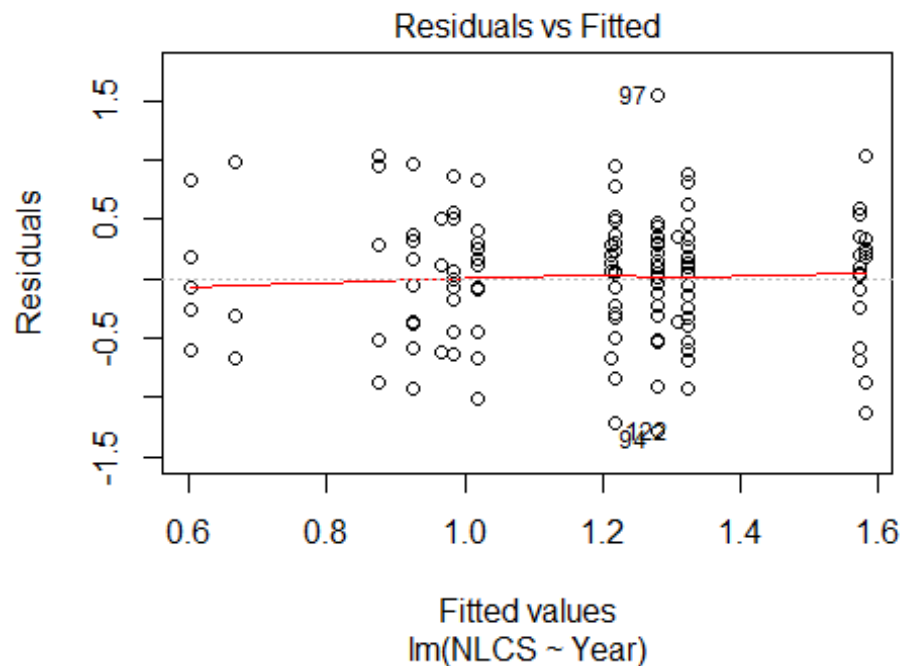


```

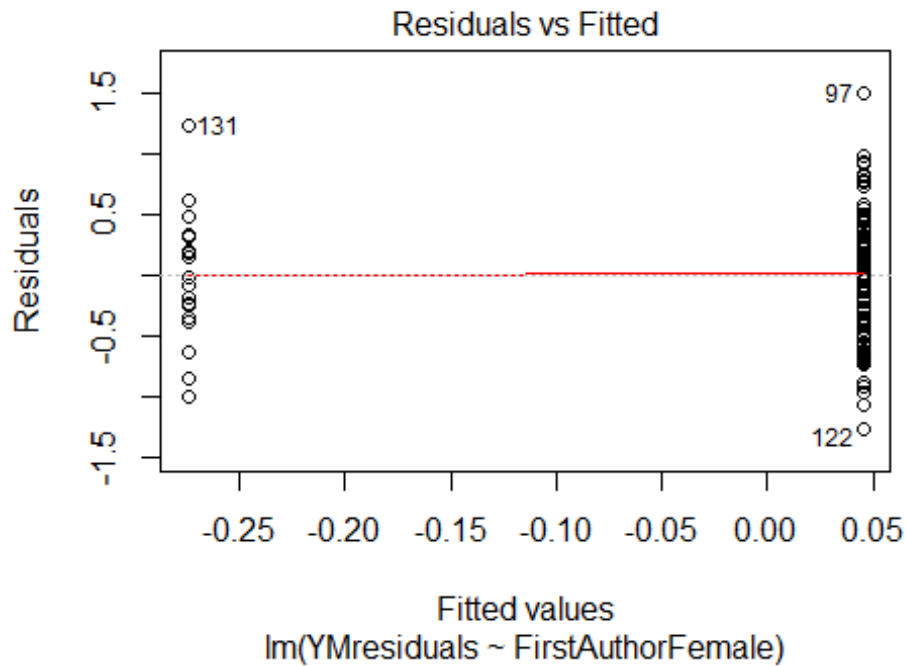
## Year2007          -6.42e-01    8.34e-02 -7.70e+00    1.0e-13 ***
## Year2008          -4.26e-01    7.19e-02 -5.93e+00    6.5e-09 ***
## Year2009          -4.19e-01    8.68e-02 -4.83e+00    1.9e-06 ***
## Year2010          -4.50e-01    8.36e-02 -5.38e+00    1.2e-07 ***
## Year2011          -3.92e-01    6.97e-02 -5.62e+00    3.5e-08 ***
## Year2012          -5.84e-01    7.79e-02 -7.50e+00    4.0e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.495
## Multiple R-squared:  0.0867, Adjusted R-squared:  0.0489
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 403 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.196  0.880  0.955  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      2.33e-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: 429"
## [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]"
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    3    3    6    2   12   18   10    6    7   23   15   18   23    9
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    3    3    6    2   12   11   10    5    6   23   14   17   22    7
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    3    3    6    2   12   11    9    4    6   21   12   17   22    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 = 7.9, df = 13, p-value = 0.9
```

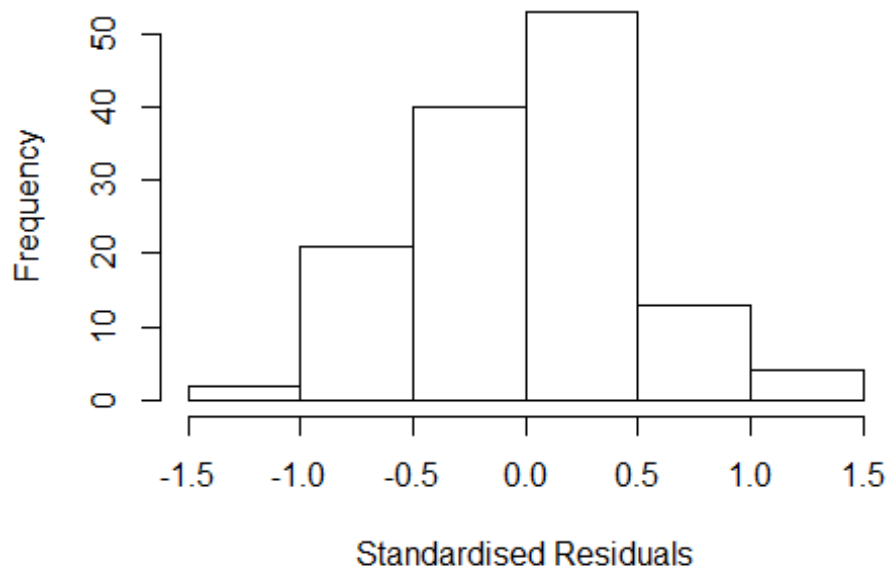


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0042, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 2.84847095691632"
## [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 2.293 1          1.514
## LastAuthorFemale  2.408 1          1.552
## UniqueAuthors    8.063 4          1.298
## Year              9.440 13         1.090
```

Residuals from first and last author and team size



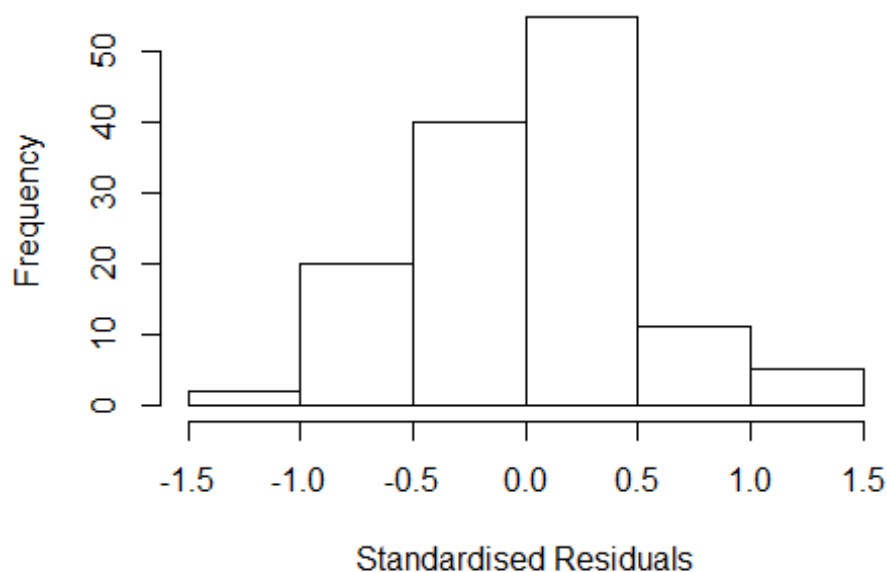
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3379 -0.3361 0.0333 0.2934 1.4971
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3833 0.4632 0.83 0.410
## FirstAuthorFemale1 0.2557 0.1537 1.66 0.099 .
## LastAuthorFemale1 0.0405 0.1533 0.26 0.792
## UniqueAuthors2 0.0696 0.1703 0.41 0.684
## UniqueAuthors3 0.0174 0.1569 0.11 0.912
## UniqueAuthors4 0.1405 0.1501 0.94 0.351
## UniqueAuthors5 -0.1114 0.1956 -0.57 0.570
## Year2000 0.3165 0.6029 0.52 0.601
## Year2001 -0.1188 0.4955 -0.24 0.811
## Year2002 0.6430 0.5741 1.12 0.265
```

```

## Year2003          0.3338      0.4904      0.68      0.497
## Year2004          0.2560      0.5044      0.51      0.613
## Year2005          0.4717      0.4868      0.97      0.335
## Year2006          0.4847      0.5000      0.97      0.334
## Year2007          0.0700      0.7683      0.09      0.928
## Year2008          0.6289      0.4872      1.29      0.199
## Year2009          0.9429      0.5154      1.83      0.070 .
## Year2010          0.6583      0.4936      1.33      0.185
## Year2011          0.6575      0.4841      1.36      0.177
## Year2012          1.1620      0.5138      2.26      0.026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.505
## Multiple R-squared:  0.29,    Adjusted R-squared:  0.171
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~ = 1. The remaining 117 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.360  0.860  0.956   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      7.52e-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 2.020 1      1.421
## LastAuthorFemale 1.881 1      1.371
## Year              1.881 13      1.025

```

Residuals from first and last author



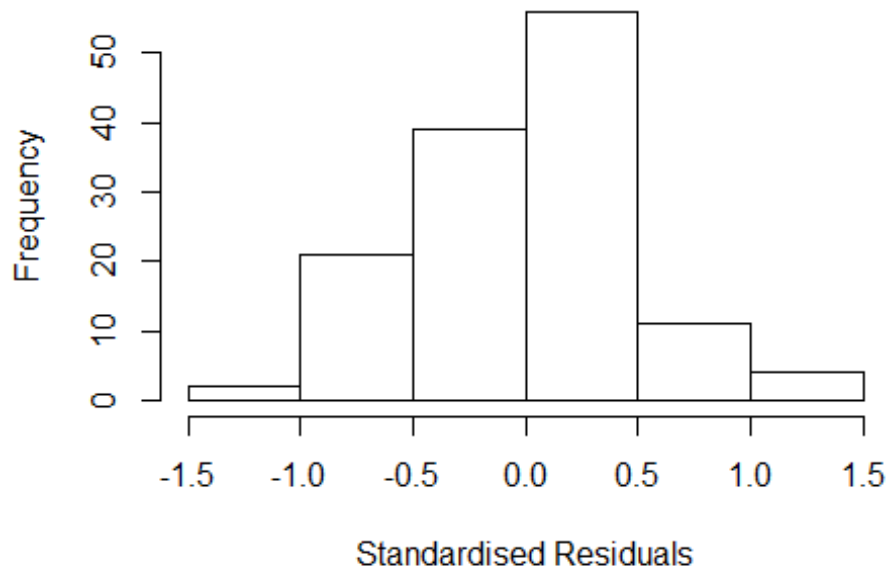
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3459 -0.3485 0.0528 0.3385 1.4495
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3483 0.4907 0.71 0.479
## FirstAuthorFemale1 0.2401 0.1446 1.66 0.100 .
## LastAuthorFemale1 0.0851 0.1443 0.59 0.556
## Year2000 0.3369 0.6159 0.55 0.585
## Year2001 -0.0982 0.5224 -0.19 0.851
## Year2002 0.6801 0.6075 1.12 0.265
## Year2003 0.4027 0.5133 0.78 0.434
## Year2004 0.3129 0.5236 0.60 0.551
## Year2005 0.4697 0.5010 0.94 0.350
## Year2006 0.5520 0.5367 1.03 0.306
## Year2007 0.1188 0.8214 0.14 0.885
## Year2008 0.7000 0.5017 1.40 0.166
```

```

## Year2009          1.0025      0.5170      1.94      0.055 .
## Year2010          0.6724      0.5097      1.32      0.190
## Year2011          0.6821      0.5096      1.34      0.183
## Year2012          1.2901      0.5085      2.54      0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.475
## Multiple R-squared:  0.287, Adjusted R-squared:  0.196
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.332  0.854   0.950   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      7.52e-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 1.322 1          1.150
## Year              1.322 13          1.011

```

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
## -1.3346 -0.3486 0.0823 0.3163 1.4601
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3696 0.4532 0.82 0.416
## FirstAuthorFemale1 0.2851 0.1179 2.42 0.017 *
## Year2000 0.3290 0.5686 0.58 0.564
## Year2001 -0.0796 0.4997 -0.16 0.874
## Year2002 0.6562 0.5435 1.21 0.230
## Year2003 0.4266 0.4902 0.87 0.386
## Year2004 0.3077 0.4905 0.63 0.532
## Year2005 0.4845 0.4767 1.02 0.312
## Year2006 0.5709 0.5125 1.11 0.268
## Year2007 0.1366 0.8137 0.17 0.867
## Year2008 0.7081 0.4735 1.50 0.137
## Year2009 1.0246 0.4932 2.08 0.040 *
```



```

## Year2010          0.6798      0.4826      1.41      0.162
## Year2011          0.6849      0.4794      1.43      0.156
## Year2012          1.3087      0.4854      2.70      0.008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.285, Adjusted R-squared:  0.201
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.323  0.855  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      7.52e-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 1.266 1          1.125
## Year            1.266 13          1.009
##
## [1] "List of 0 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
## -1.2953 -0.3622  0.0436  0.3360  1.4577
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)

```

```

## (Intercept)          0.4484      0.5736      0.78      0.436
## LastAuthorFemale1    0.1866      0.1183      1.58      0.118
## Year2000              0.4024      0.6668      0.60      0.547
## Year2001             -0.0551      0.5696     -0.10      0.923
## Year2002              0.7693      0.6728      1.14      0.255
## Year2003              0.4091      0.5651      0.72      0.471
## Year2004              0.3578      0.5754      0.62      0.535
## Year2005              0.4536      0.5664      0.80      0.425
## Year2006              0.5855      0.5818      1.01      0.316
## Year2007              0.1840      0.7705      0.24      0.812
## Year2008              0.7302      0.5548      1.32      0.191
## Year2009              0.9828      0.5684      1.73      0.086 .
## Year2010              0.6602      0.5650      1.17      0.245
## Year2011              0.7282      0.5587      1.30      0.195
## Year2012              1.3320      0.5570      2.39      0.018 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.264, Adjusted R-squared:  0.176
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 125 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.411  0.878  0.951  0.916  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          7.52e-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: 133"
## [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]"

```

```

##
## 1998 1999 2000 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    2    1    1    6    2    1    1    4   12    6    9   10   13    5
##
## 1998 1999 2000 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    0    4    2    1    1    4   11    6    9   10   12    5
##
## 1998 1999 2000 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    0    0    3    2    1    1    4   10    5    8    8   10    5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.9646987343107"
## [1] "Male first author team size 2018 geometric mean: 3.46410161513775"

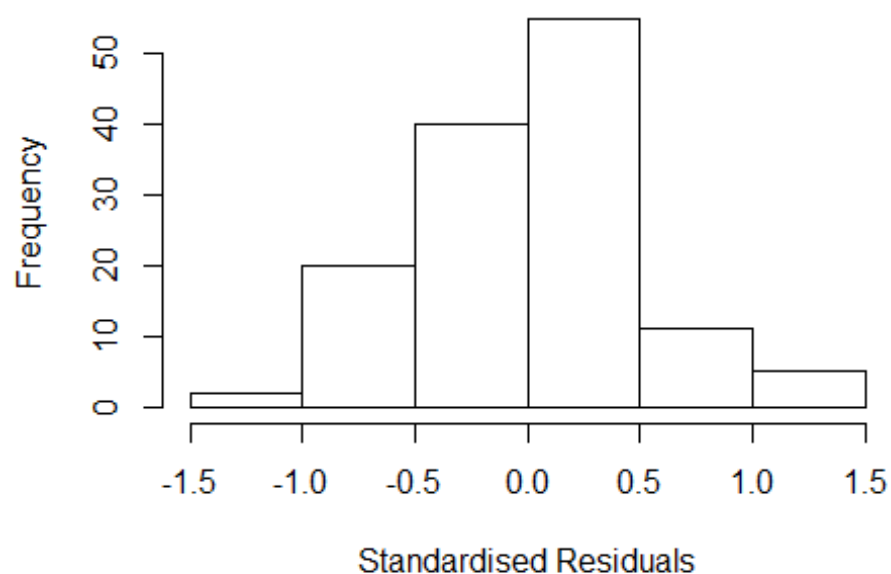
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.86388853784137"
## [1] "Male last author team size 2018 geometric mean: 3.46410161513775"

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

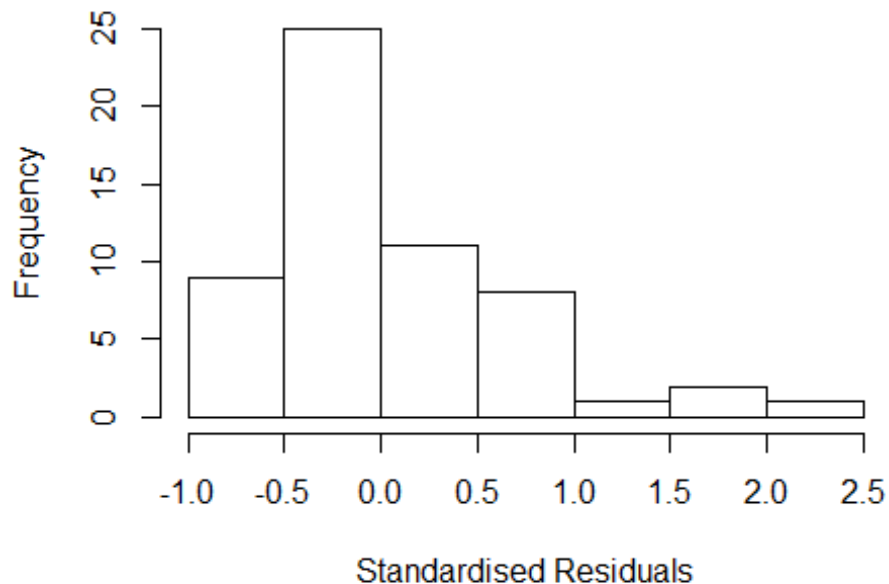
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, 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 9.807e+13  1      9.903e+06
## LastAuthorFemale  1.051e+14  1      1.025e+07
## UniqueAuthors    8.268e+15  4      9.765e+01
## Year              8.406e+16 10      7.018e+00
```

Residuals from first and last author and team size



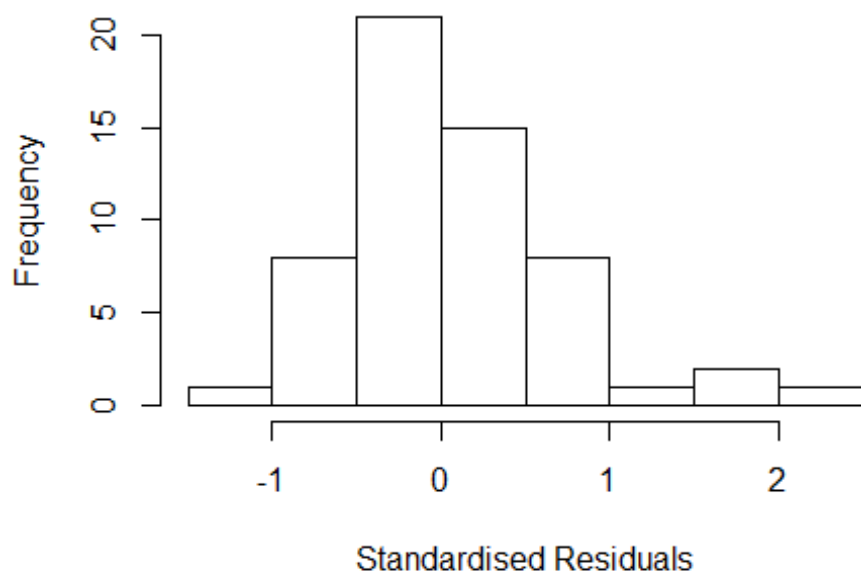
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9130 -0.2855 -0.0435 0.4569 2.1424
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5618 0.6040 0.93 0.36
## FirstAuthorFemale1 0.2478 0.2077 1.19 0.24
## LastAuthorFemale1 -0.2040 0.2150 -0.95 0.35
## UniqueAuthors2 -0.0197 0.3902 -0.05 0.96
## UniqueAuthors3 0.3791 0.2712 1.40 0.17
## UniqueAuthors4 0.1673 0.3497 0.48 0.63
## UniqueAuthors5 0.1573 0.4084 0.39 0.70
## Year2002 0.2545 0.7473 0.34 0.74
## Year2004 0.7632 0.6040 1.26 0.21
## Year2005 0.7733 0.5942 1.30 0.20
```

```

## Year2006          0.5827      0.6203      0.94      0.35
## Year2007         -0.0724      0.6655     -0.11      0.91
## Year2008          0.7528      0.8342      0.90      0.37
## Year2009          0.3999      0.5043      0.79      0.43
## Year2010          0.4055      0.6904      0.59      0.56
## Year2011          0.8282      0.6085      1.36      0.18
## Year2012          0.1146      0.7084      0.16      0.87
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.339, Adjusted R-squared:  0.0743
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 49 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0705 0.8690 0.9310 0.8770 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      1.75e-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 9.604e+14 1      3.099e+07
## LastAuthorFemale 1.376e+15 1      3.710e+07
## Year              4.483e+15 10      6.061e+00

```

Residuals from first and last author



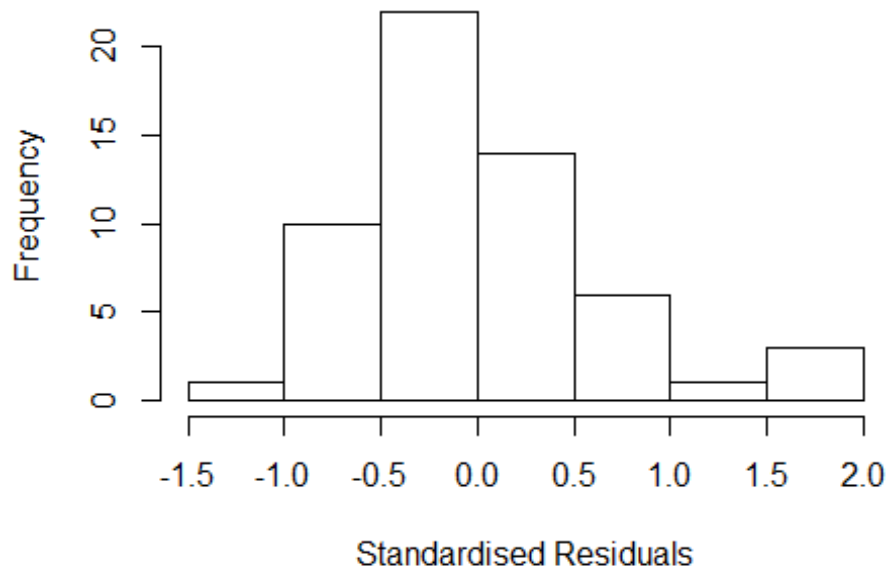
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.03e+00 -3.48e-01 -4.44e-16 4.03e-01 2.05e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.57887 0.54072 1.07 0.290
## FirstAuthorFemale1 0.28864 0.16362 1.76 0.085 .
## LastAuthorFemale1 -0.25717 0.19586 -1.31 0.196
## Year2002 0.31240 0.52118 0.60 0.552
## Year2004 0.74613 0.54072 1.38 0.175
## Year2005 1.14766 0.46432 2.47 0.017 *
## Year2006 0.69431 0.55700 1.25 0.219
## Year2007 -0.00108 0.63356 0.00 0.999
## Year2008 0.88746 0.60157 1.48 0.147
## Year2009 0.49180 0.49129 1.00 0.322
## Year2010 0.42752 0.49094 0.87 0.389
## Year2011 0.92638 0.50690 1.83 0.074 .
```

```

## Year2012          0.33987    0.58223    0.58    0.562
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.293, Adjusted R-squared:  0.1
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 50 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.153  0.885  0.948  0.883  0.992  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      1.75e-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 9.286e+13 1      9.637e+06
## Year              9.286e+13 10      4.993e+00

```


Residuals from first author



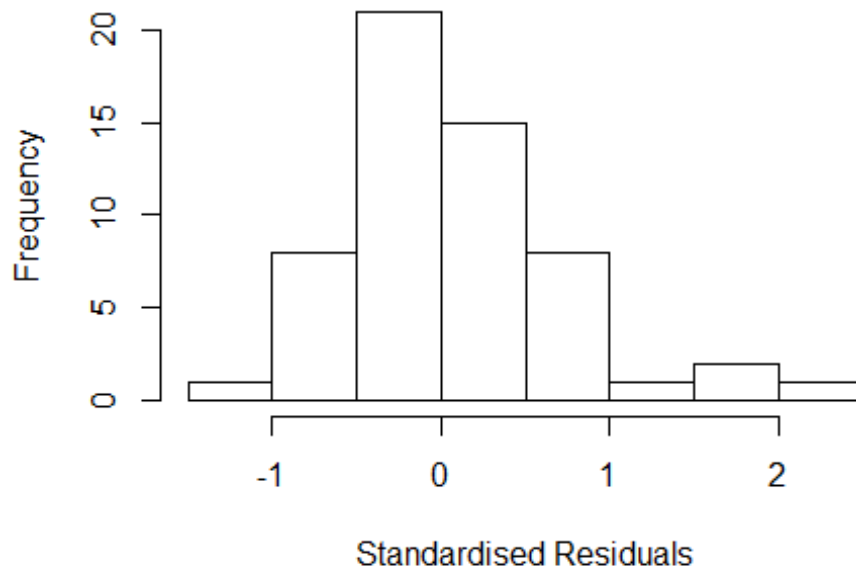
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1024 -0.3147 -0.0411 0.4231 1.8826
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.392 0.503 0.78 0.439
## FirstAuthorFemale1 0.228 0.204 1.11 0.271
## Year2002 0.401 0.563 0.71 0.480
## Year2004 0.933 0.503 1.85 0.070 .
## Year2005 1.138 0.447 2.54 0.014 *
## Year2006 0.765 0.538 1.42 0.162
## Year2007 0.197 0.583 0.34 0.737
## Year2008 0.931 0.575 1.62 0.113
## Year2009 0.562 0.479 1.17 0.247
## Year2010 0.549 0.473 1.16 0.252
## Year2011 0.995 0.497 2.00 0.052 .
## Year2012 0.482 0.550 0.88 0.385
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.216, Adjusted R-squared:  0.0246
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 48 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.315  0.865  0.950  0.892  0.983  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           1.75e-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 7.011e+14 1           2.648e+07
## Year           7.011e+14 10           5.524e+00

```

Residuals from last author



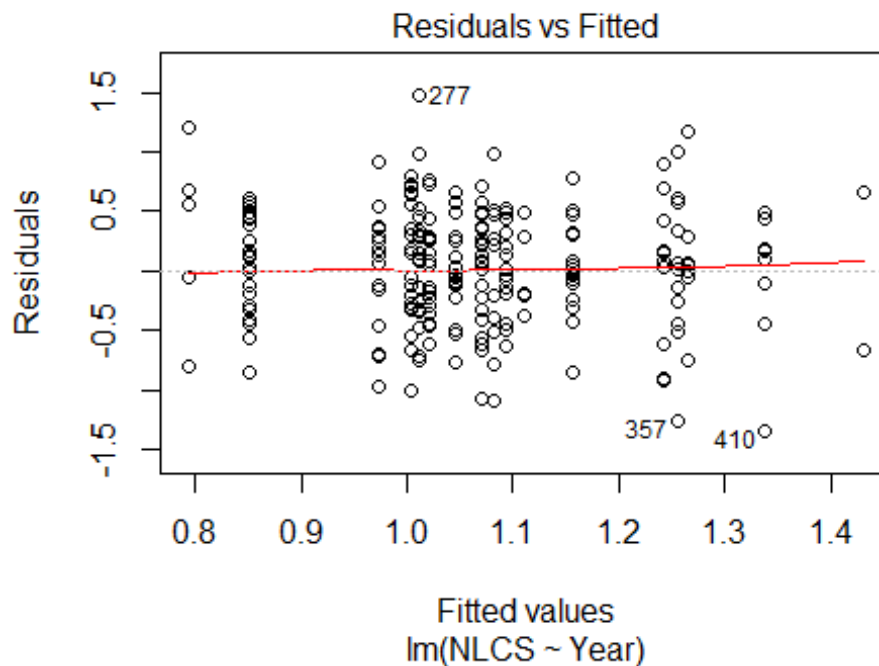
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14e+00 -3.69e-01 -3.33e-16 3.51e-01 2.16e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.822 0.500 1.64 0.107
## LastAuthorFemale1 -0.215 0.214 -1.00 0.320
## Year2002 0.193 0.585 0.33 0.743
## Year2004 0.503 0.500 1.01 0.320
## Year2005 1.151 0.469 2.45 0.018 *
## Year2006 0.658 0.530 1.24 0.221
## Year2007 -0.111 0.624 -0.18 0.860
## Year2008 0.743 0.644 1.15 0.255
## Year2009 0.482 0.502 0.96 0.342
## Year2010 0.338 0.488 0.69 0.492
## Year2011 0.791 0.494 1.60 0.117
## Year2012 0.233 0.572 0.41 0.686
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.266, Adjusted R-squared:  0.0866
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 51 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0836 0.8610 0.9510 0.8770 0.9830 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           1.75e-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:  57"
## [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
##    13    8   17   21   28   30   26   20   23   21   33   30   34   13   12
## 2011 2012
##    15   13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     7    2   10   18    5   15   21   18   16   15   26   23   24   11   11
## 2011 2012
##    11    9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     6    2   10   15    5   14   18   16   15   15   26   20   21    8   10
## 2011 2012
##     8    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 = 20, df = 16, p-value = 0.2
```



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

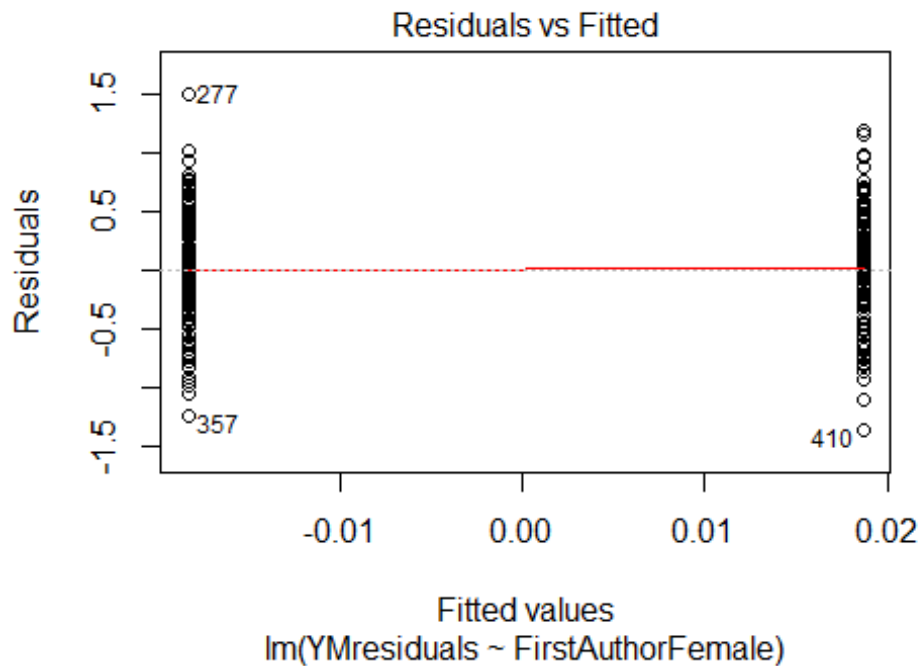
## [1] "Female first author team size 2018 geometric mean: 3.98375933805648"
## [1] "Male first author team size 2018 geometric mean: 2.57283925640405"

## 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 = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
```

```
## [1] "Female last author team size 2018 geometric mean: 3.16644697529495"
## [1] "Male last author team size 2018 geometric mean: 3.66210672754074"

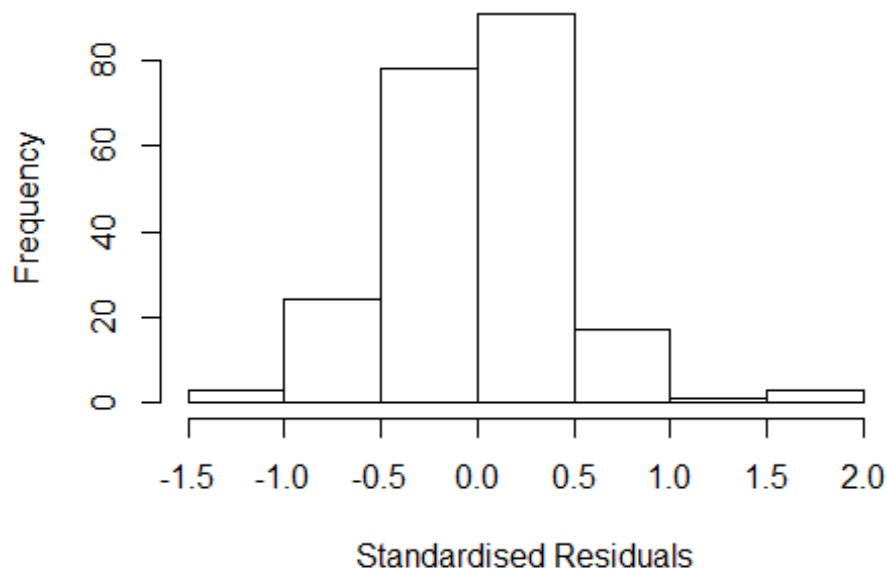
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 71, 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	1.972	1	1.404
LastAuthorFemale	2.923	1	1.710
UniqueAuthors	8.636	4	1.309
Year	8.949	16	1.071

Residuals from first and last author and team size



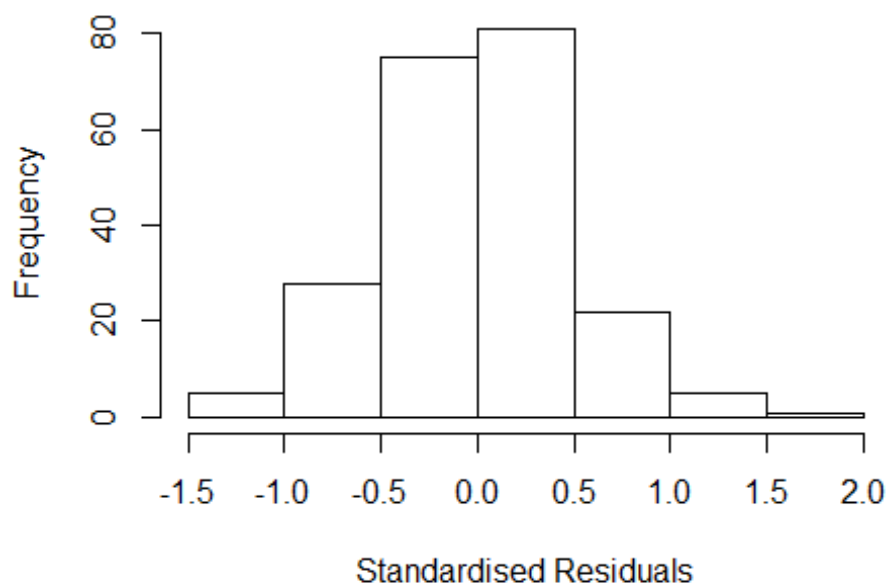
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.232 -0.247 0.022 0.264 1.999
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.70154 0.27069 2.59 0.010 *
## FirstAuthorFemale1 0.00176 0.07058 0.02 0.980
## LastAuthorFemale1 -0.07725 0.09095 -0.85 0.397
## UniqueAuthors2 0.33469 0.14625 2.29 0.023 *
## UniqueAuthors3 0.52298 0.10675 4.90 2.0e-06 ***
## UniqueAuthors4 0.55602 0.10646 5.22 4.5e-07 ***
## UniqueAuthors5 0.63986 0.10903 5.87 1.9e-08 ***
## Year1997 0.22859 1.25623 0.18 0.856
## Year1998 0.15744 0.33692 0.47 0.641
## Year1999 -0.06512 0.26493 -0.25 0.806
```

```

## Year2000      -0.01567    0.33697   -0.05    0.963
## Year2001      -0.11393    0.27552   -0.41    0.680
## Year2002       0.07223    0.26246    0.28    0.783
## Year2003      -0.06237    0.26935   -0.23    0.817
## Year2004      -0.05258    0.26568   -0.20    0.843
## Year2005      -0.06719    0.28674   -0.23    0.815
## Year2006      -0.09618    0.25450   -0.38    0.706
## Year2007      -0.13857    0.26805   -0.52    0.606
## Year2008      -0.03694    0.26769   -0.14    0.890
## Year2009       0.07166    0.33594    0.21    0.831
## Year2010      -0.10963    0.28823   -0.38    0.704
## Year2011       0.25805    0.34424    0.75    0.454
## Year2012      -0.04958    0.36640   -0.14    0.893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.335, Adjusted R-squared:  0.26
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## observation 143 is an outlier with |weight| = 0 ( < 0.00046);
## 26 weights are ~= 1. The remaining 190 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0114 0.8720 0.9530 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          4.61e-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 1.577 1 1.256
## LastAuthorFemale 1.776 1 1.333
## Year 2.026 16 1.022

```


Residuals from first and last author



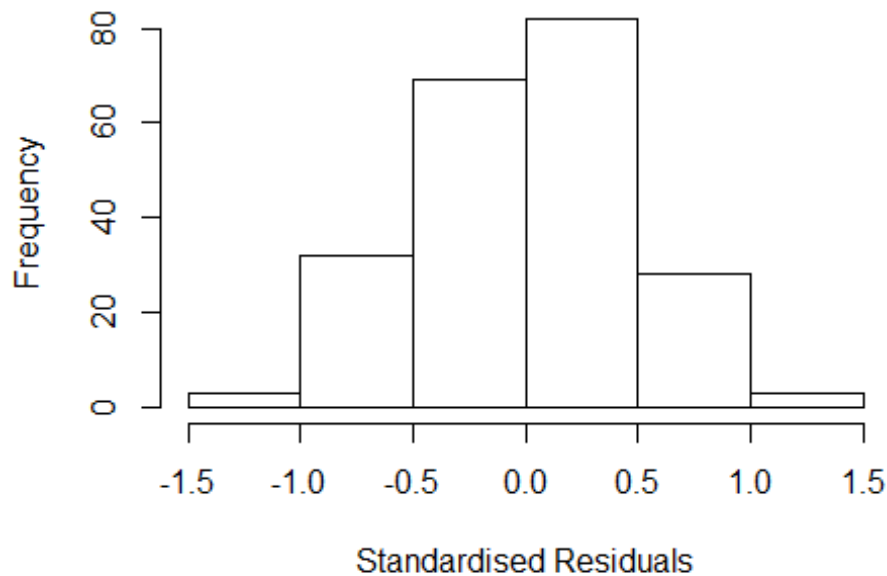
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2035 -0.3067 0.0011 0.3009 1.6484
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1061 0.6189 1.79 0.0755 .
## FirstAuthorFemale1 0.0302 0.0782 0.39 0.7001
## LastAuthorFemale1 -0.2563 0.0917 -2.79 0.0057 **
## Year1997 0.4531 1.5747 0.29 0.7739
## Year1998 0.2438 0.6645 0.37 0.7141
## Year1999 0.0774 0.6229 0.12 0.9012
## Year2000 0.0925 0.6248 0.15 0.8824
## Year2001 0.0172 0.6226 0.03 0.9780
## Year2002 0.0522 0.6178 0.08 0.9327
## Year2003 -0.0218 0.6266 -0.03 0.9723
## Year2004 0.0146 0.6249 0.02 0.9814
## Year2005 -0.0274 0.6306 -0.04 0.9654
```

```

## Year2006          -0.1432      0.6153   -0.23   0.8162
## Year2007          -0.0423      0.6346   -0.07   0.9469
## Year2008          -0.0182      0.6250   -0.03   0.9768
## Year2009           0.3020      0.6752    0.45   0.6552
## Year2010          -0.0976      0.6474   -0.15   0.8803
## Year2011           0.3537      0.6355    0.56   0.5785
## Year2012           0.1632      0.6424    0.25   0.7997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.117, Adjusted R-squared:  0.0372
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 191 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.179  0.870   0.942   0.898   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      4.61e-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 1.278 1         1.131
## Year              1.278 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
## -1.4074 -0.3134 0.0224 0.3622 1.4763
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9641 0.4697 2.05 0.041 *
## FirstAuthorFemale1 -0.0550 0.0737 -0.75 0.457
## Year1997 0.4669 0.8608 0.54 0.588
## Year1998 0.3350 0.5176 0.65 0.518
## Year1999 0.1548 0.4785 0.32 0.747
## Year2000 0.1629 0.4911 0.33 0.740
## Year2001 0.1301 0.4791 0.27 0.786
## Year2002 0.1412 0.4783 0.30 0.768
## Year2003 0.0936 0.4813 0.19 0.846
## Year2004 0.1524 0.4783 0.32 0.750
## Year2005 0.0726 0.4886 0.15 0.882
## Year2006 -0.0710 0.4758 -0.15 0.882
```

```

## Year2007          0.1006      0.4886      0.21      0.837
## Year2008          0.0792      0.4869      0.16      0.871
## Year2009          0.4058      0.5309      0.76      0.446
## Year2010          0.0729      0.5149      0.14      0.888
## Year2011          0.4433      0.4964      0.89      0.373
## Year2012          0.2557      0.5089      0.50      0.616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0699, Adjusted R-squared:  -0.0096
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.365  0.877  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      4.61e-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 1.53  1      1.237
## Year              1.53 16      1.013

## [1] "List of 0 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
## -1.22876 -0.30113  0.00741  0.31428  1.65236

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11302    0.62007    1.79  0.0742 .
## LastAuthorFemale1 -0.24595    0.08498   -2.89  0.0042 **
## Year1997        0.44096    1.56096    0.28  0.7779
## Year1998        0.24925    0.66405    0.38  0.7078
## Year1999        0.07791    0.62388    0.12  0.9007
## Year2000        0.09311    0.62482    0.15  0.8817
## Year2001        0.01663    0.62330    0.03  0.9787
## Year2002        0.05474    0.61887    0.09  0.9296
## Year2003       -0.01318    0.62581   -0.02  0.9832
## Year2004        0.02261    0.62555    0.04  0.9712
## Year2005       -0.01547    0.63065   -0.02  0.9804
## Year2006       -0.13330    0.61552   -0.22  0.8288
## Year2007       -0.03343    0.63564   -0.05  0.9581
## Year2008       -0.00716    0.62538   -0.01  0.9909
## Year2009        0.29521    0.67320    0.44  0.6615
## Year2010       -0.09141    0.65037   -0.14  0.8884
## Year2011        0.36170    0.63519    0.57  0.5697
## Year2012        0.15776    0.64442    0.24  0.8069
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.118, Adjusted R-squared:  0.0421
## Convergence in 31 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 190 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.171  0.863   0.941   0.896   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      4.61e-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: 217"
## [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]"
##
## 1997 2001 2002 2005 2007 2008 2009 2010 2011 2012
##    1    1    1    1    2    4    4    2    1    5
##
## 1997 2001 2002 2005 2007 2008 2009 2010 2011 2012
##    0    1    0    1    2    4    3    2    0    5
##
## 1997 2001 2002 2005 2007 2008 2009 2010 2011 2012
##    0    1    0    1    2    4    3    2    0    5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 6.30868429105981"
## [1] "Male first author team size 2018 geometric mean: 8.36660026534076"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.41506865999165"
## [1] "Male last author team size 2018 geometric mean: 7.48887238721851"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 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"

## 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"

## 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

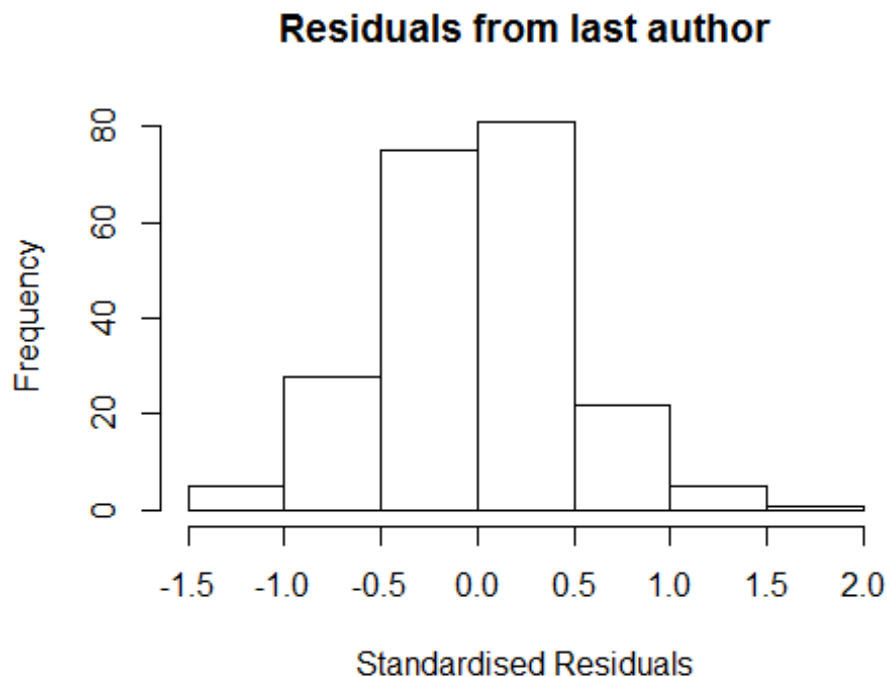
```

```
## 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

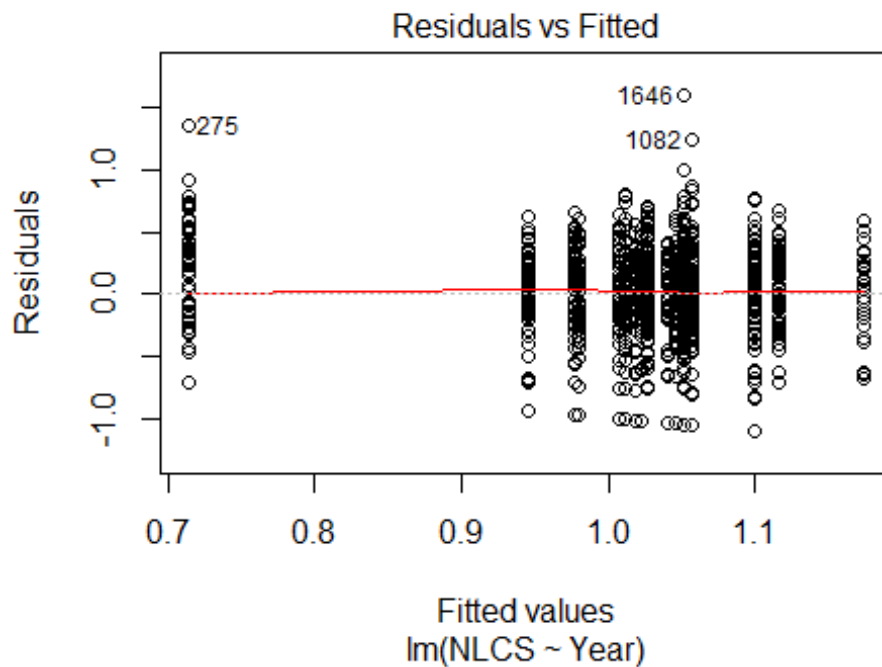
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

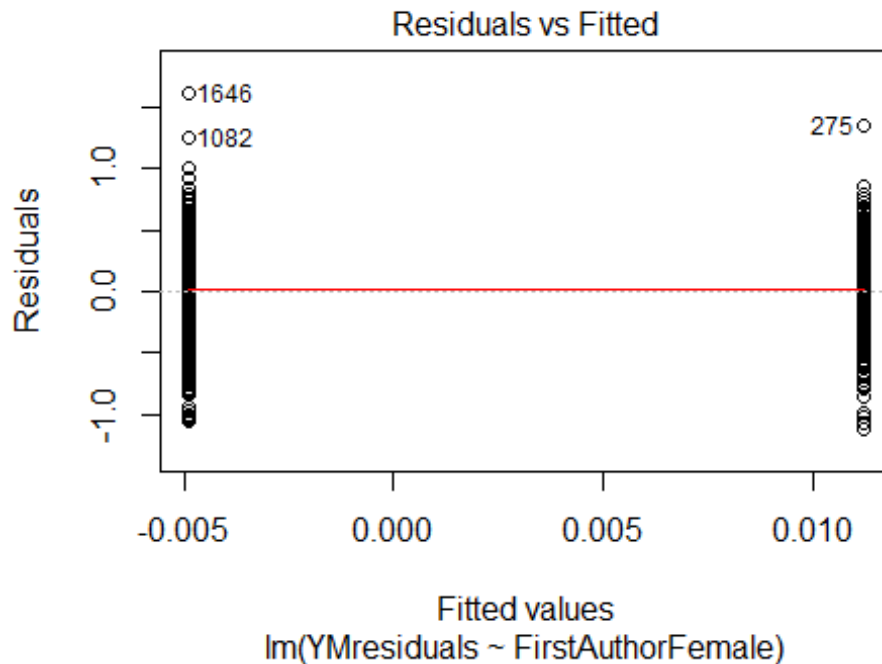


```
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 18"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3002"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    71   84   92   79   85   84   64   39   71   85  106  115  107  117  126
## 2011 2012
##  167  143
```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   53   63   72   64   55   60   46   28   47   73   79   89   82   83   86
## 2011 2012
##  126  102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   49   60   69   58   47   51   39   26   45   68   75   85   78   80   79
## 2011 2012
##  114   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 = 53, df = 16, p-value = 7e-06
```

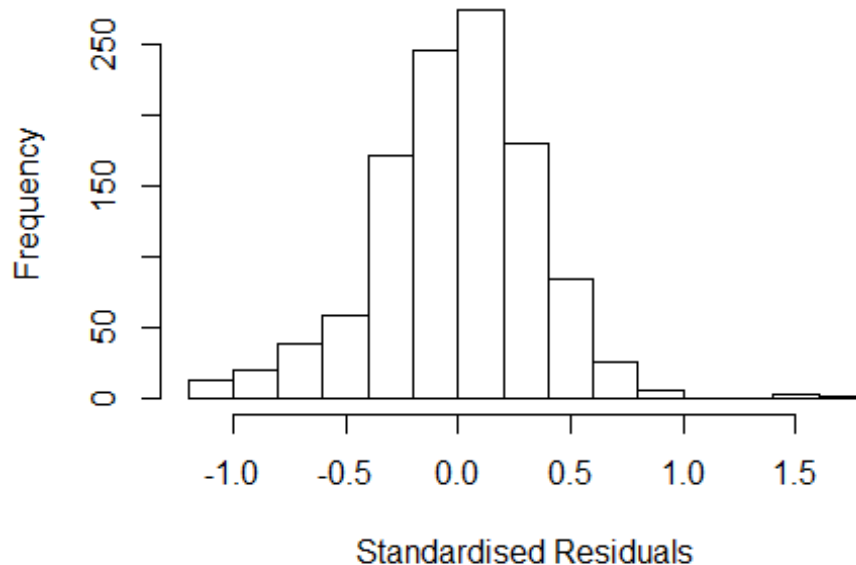


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.024, df = 1, p-value = 0.9
```

```
## [1] "Female first author team size 2018 geometric mean: 5.82672422871442"
## [1] "Male first author team size 2018 geometric mean: 4.30149993339731"
##
## 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: 4.27135709529921"
## [1] "Male last author team size 2018 geometric mean: 4.80631302207496"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 540, 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 1.105 1          1.051
## LastAuthorFemale  1.134 1          1.065
## UniqueAuthors    1.561 4          1.057
## Year              1.628 16         1.015
```

Residuals from first and last author and team size



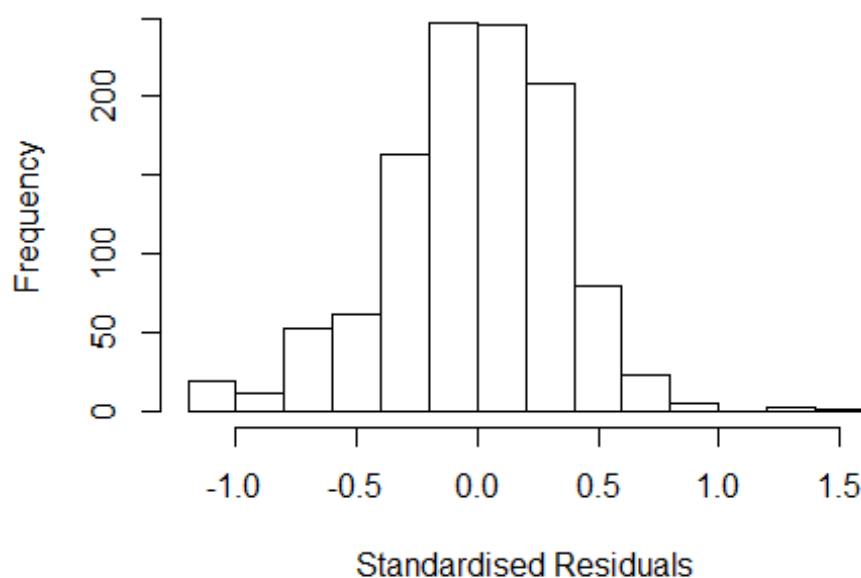
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1507 -0.2208  0.0142  0.2235  1.6963
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9293    0.0856   10.86 < 2e-16 ***
## FirstAuthorFemale1  0.0139    0.0225    0.62  0.53641
## LastAuthorFemale1  0.0226    0.0268    0.84  0.40076
## UniqueAuthors2    0.0841    0.0773    1.09  0.27671
## UniqueAuthors3    0.1482    0.0764    1.94  0.05252 .
## UniqueAuthors4    0.1490    0.0750    1.99  0.04707 *
## UniqueAuthors5    0.2520    0.0740    3.41  0.00068 ***
## Year1997         -0.0447    0.0663   -0.67  0.50026
## Year1998         -0.3647    0.0993   -3.67  0.00025 ***
## Year1999         -0.0366    0.0658   -0.56  0.57884
```

```

## Year2000          -0.0410      0.0707    -0.58  0.56266
## Year2001          -0.1030      0.0725    -1.42  0.15538
## Year2002          -0.0260      0.0715    -0.36  0.71626
## Year2003           0.0726      0.0857     0.85  0.39651
## Year2004          -0.0334      0.0610    -0.55  0.58415
## Year2005          -0.1223      0.0617    -1.98  0.04784 *
## Year2006          -0.0493      0.0570    -0.86  0.38726
## Year2007          -0.0576      0.0605    -0.95  0.34126
## Year2008           0.0185      0.0628     0.29  0.76904
## Year2009          -0.0175      0.0587    -0.30  0.76549
## Year2010          -0.0792      0.0598    -1.32  0.18557
## Year2011          -0.0859      0.0581    -1.48  0.13971
## Year2012          -0.0446      0.0601    -0.74  0.45827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.323
## Multiple R-squared:  0.116, Adjusted R-squared:  0.0982
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 888 is an outlier with |weight| = 0 ( < 8.9e-05);
## 81 weights are ~= 1. The remaining 1037 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0118 0.8770 0.9500 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          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 1.085 1          1.042
## LastAuthorFemale  1.088 1          1.043
## Year              1.158 16          1.005

```

Residuals from first and last author



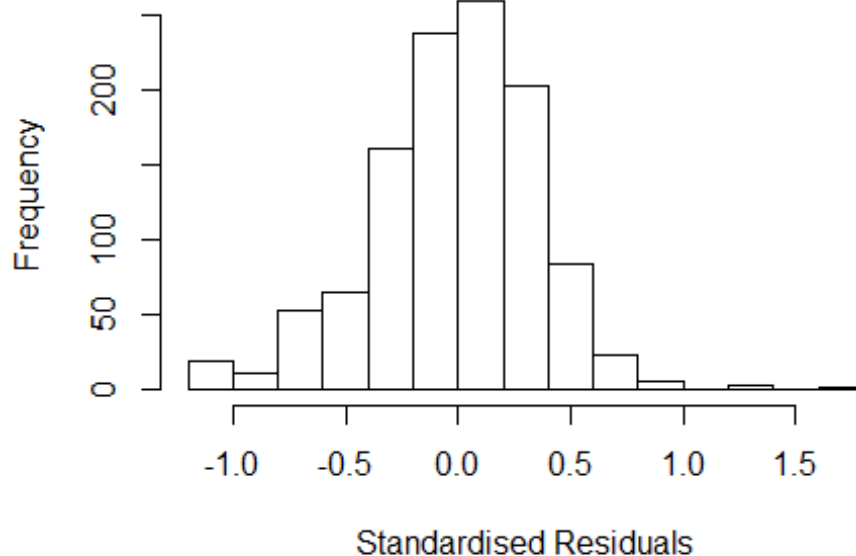
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17870 -0.21812 0.00514 0.22824 1.58468
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05632 0.04340 24.34 <2e-16 ***
## FirstAuthorFemale1 0.01874 0.02281 0.82 0.4115
## LastAuthorFemale1 0.04010 0.02790 1.44 0.1509
## Year1997 -0.03936 0.06299 -0.62 0.5322
## Year1998 -0.36097 0.10333 -3.49 0.0005 ***
## Year1999 -0.02112 0.06313 -0.33 0.7381
## Year2000 -0.01269 0.06888 -0.18 0.8539
## Year2001 -0.05812 0.06951 -0.84 0.4033
## Year2002 -0.00285 0.06538 -0.04 0.9652
## Year2003 0.12746 0.08532 1.49 0.1355
## Year2004 0.00810 0.05910 0.14 0.8910
## Year2005 -0.10039 0.05886 -1.71 0.0884 .
```

```

## Year2006          0.00058    0.05211    0.01    0.9911
## Year2007         -0.00600    0.05688   -0.11    0.9160
## Year2008          0.06354    0.05799    1.10    0.2735
## Year2009          0.03784    0.05407    0.70    0.4841
## Year2010         -0.02810    0.05657   -0.50    0.6194
## Year2011         -0.03847    0.05257   -0.73    0.4644
## Year2012          0.00622    0.05612    0.11    0.9117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0705, Adjusted R-squared:  0.0553
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 888 is an outlier with |weight| = 0 ( < 8.9e-05);
## 94 weights are ~= 1. The remaining 1024 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0538 0.8720 0.9510 0.8940 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      8.94e-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 1.077 1          1.038
## Year              1.077 16          1.002

```

Residuals from first author



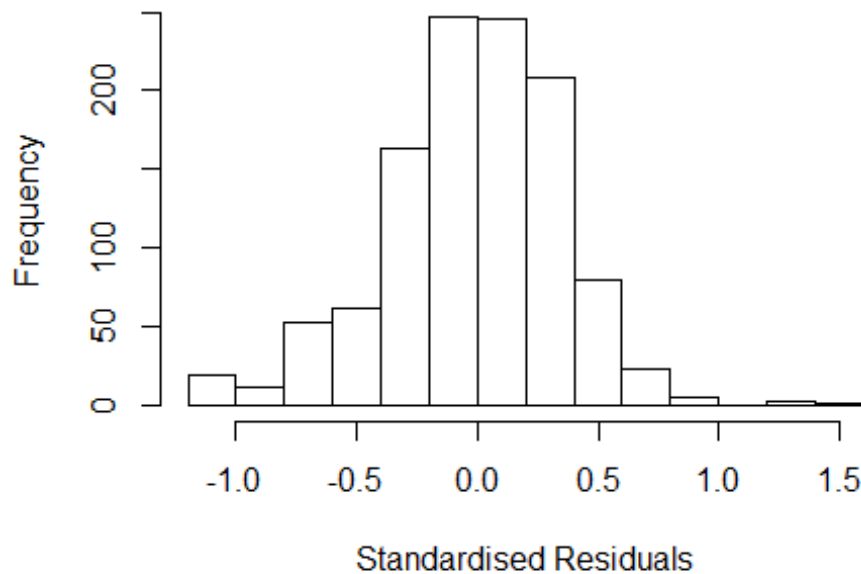
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1419 -0.2212  0.0137  0.2283  1.6182
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.059742   0.043808   24.19  < 2e-16 ***
## FirstAuthorFemale1 0.018499   0.022841    0.81  0.41817
## Year1997       -0.039751   0.063345   -0.63  0.53044
## Year1998       -0.360505   0.104635   -3.45  0.00059 ***
## Year1999       -0.017524   0.063216   -0.28  0.78167
## Year2000       -0.008494   0.068959   -0.12  0.90199
## Year2001       -0.056118   0.070102   -0.80  0.42359
## Year2002        0.001996   0.066089    0.03  0.97591
## Year2003        0.134026   0.084660    1.58  0.11368
## Year2004        0.016594   0.059033    0.28  0.77869
## Year2005       -0.097875   0.059181   -1.65  0.09845 .
## Year2006       -0.000491   0.052478   -0.01  0.99254
```

```

## Year2007      -0.003565    0.057347   -0.06   0.95044
## Year2008      0.063662    0.058465    1.09   0.27644
## Year2009      0.043561    0.054309    0.80   0.42267
## Year2010     -0.024942    0.056702   -0.44   0.66011
## Year2011     -0.033283    0.052825   -0.63   0.52879
## Year2012      0.009540    0.056345    0.17   0.86559
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.331
## Multiple R-squared:  0.0684, Adjusted R-squared:  0.0541
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 888 is an outlier with |weight| = 0 ( < 8.9e-05);
## 80 weights are ~= 1. The remaining 1038 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0581 0.8720 0.9530 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      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 1.075 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
## -1.16529 -0.22230 0.00912 0.22507 1.57861
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06288 0.04262 24.94 < 2e-16 ***
## LastAuthorFemale1 0.03990 0.02784 1.43 0.15208
## Year1997 -0.04136 0.06302 -0.66 0.51180
## Year1998 -0.36511 0.10367 -3.52 0.00045 ***
## Year1999 -0.02480 0.06313 -0.39 0.69449
## Year2000 -0.01351 0.06896 -0.20 0.84474
## Year2001 -0.05735 0.06955 -0.82 0.40983
## Year2002 -0.00610 0.06515 -0.09 0.92542
## Year2003 0.12829 0.08596 1.49 0.13586
## Year2004 0.01043 0.05882 0.18 0.85935
## Year2005 -0.10003 0.05905 -1.69 0.09055 .
## Year2006 0.00231 0.05213 0.04 0.96472
```

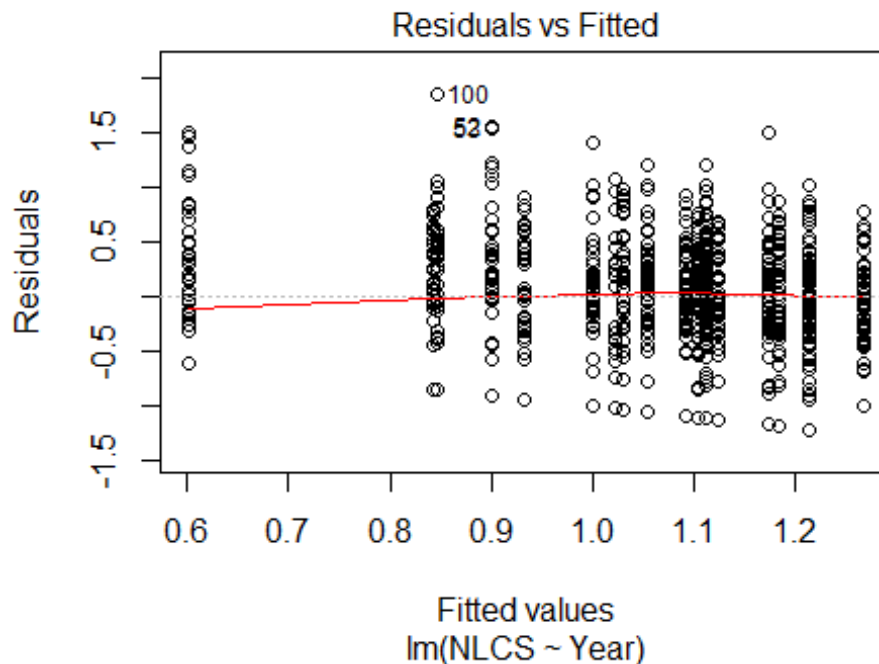


```

## Year2007      -0.00765    0.05670   -0.13  0.89267
## Year2008      0.06250    0.05799    1.08  0.28130
## Year2009      0.03716    0.05405    0.69  0.49193
## Year2010     -0.02840    0.05641   -0.50  0.61474
## Year2011     -0.03833    0.05260   -0.73  0.46637
## Year2012      0.00476    0.05620    0.08  0.93251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0697, Adjusted R-squared:  0.0553
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 888 is an outlier with |weight| = 0 ( < 8.9e-05);
## 89 weights are ~= 1. The remaining 1029 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0451 0.8710 0.9510 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      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 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
##   68   65   92   56   77   73   56   39   61   51   65   77   90  105  109
## 2011 2012
##  139  116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   54   49   65   37   49   47   37   22   47   41   51   54   59   73   70

```

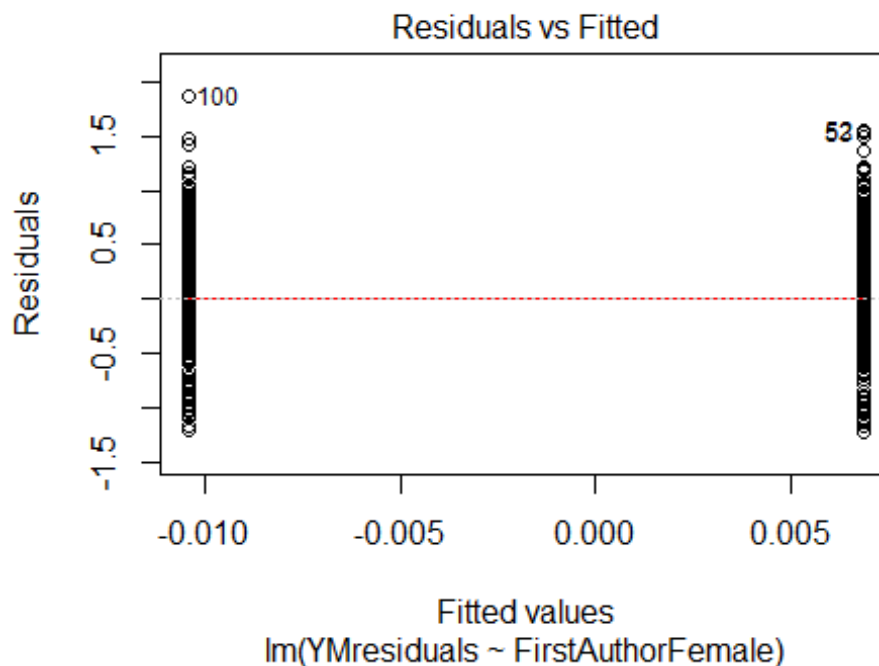
```
## 2011 2012
## 105 75
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 43 60 33 44 41 33 19 42 35 47 50 55 69 64
## 2011 2012
## 97 68
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## 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 = 4.8, df = 1, p-value = 0.03
## [1] "Female first author team size 2018 geometric mean: 4.19702773162309"
## [1] "Male first author team size 2018 geometric mean: 4.60090474723739"
## 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 = 670, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.94084322192167"
## [1] "Male last author team size 2018 geometric mean: 4.75024150753064"

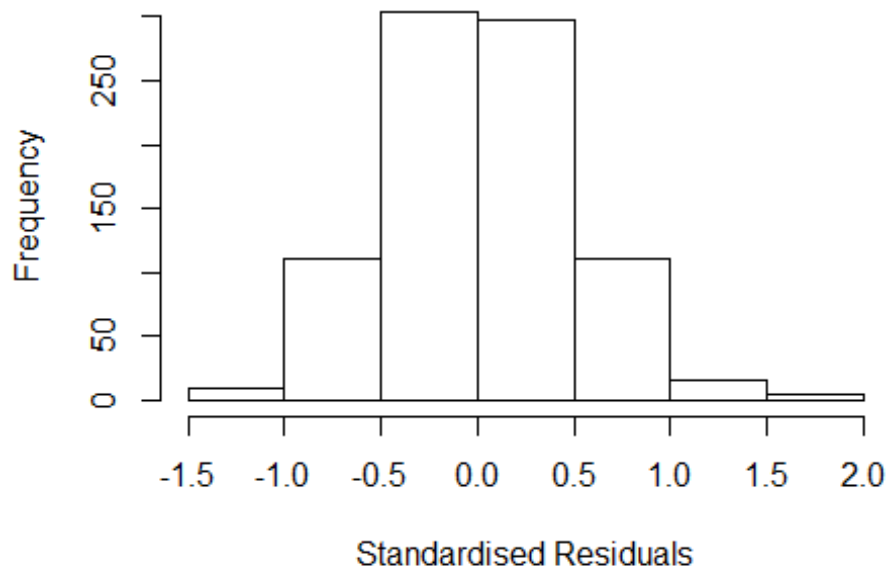
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 570, 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	1.210	1	1.100
LastAuthorFemale	1.211	1	1.100
UniqueAuthors	1.949	4	1.087
Year	2.274	16	1.026

Residuals from first and last author and team size



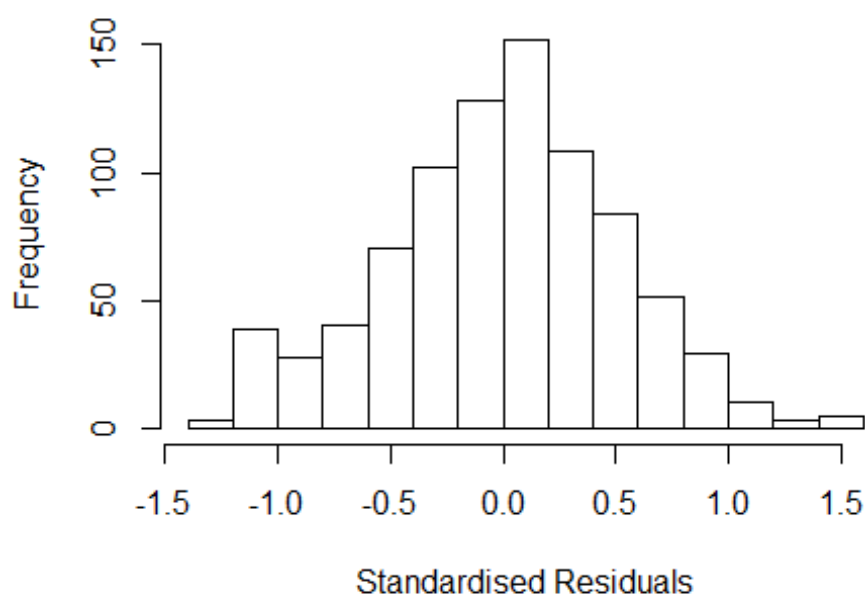
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24443 -0.33611  0.00101  0.31536  1.97247
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.58275    0.11747   4.96 8.5e-07 ***
## FirstAuthorFemale1 0.00025    0.03763   0.01  0.9947
## LastAuthorFemale1 -0.07687    0.04165  -1.85  0.0653 .
## UniqueAuthors2    0.26605    0.08276   3.21  0.0014 **
## UniqueAuthors3    0.53549    0.07691   6.96 6.8e-12 ***
## UniqueAuthors4    0.64025    0.07432   8.61 < 2e-16 ***
## UniqueAuthors5    0.58433    0.07086   8.25 6.4e-16 ***
## Year1997         -0.07855    0.13986  -0.56  0.5745
## Year1998         -0.39951    0.12732  -3.14  0.0018 **
## Year1999          0.06443    0.13564   0.48  0.6349
```

```

## Year2000      -0.03842    0.13905   -0.28    0.7824
## Year2001      -0.09294    0.14403   -0.65    0.5189
## Year2002      -0.03660    0.14277   -0.26    0.7978
## Year2003      -0.16730    0.17668   -0.95    0.3439
## Year2004       0.09681    0.12465    0.78    0.4376
## Year2005       0.04648    0.13242    0.35    0.7257
## Year2006       0.11977    0.12612    0.95    0.3425
## Year2007       0.17846    0.12180    1.47    0.1432
## Year2008       0.05401    0.11835    0.46    0.6483
## Year2009       0.05912    0.11724    0.50    0.6142
## Year2010       0.03893    0.11926    0.33    0.7442
## Year2011       0.04976    0.11745    0.42    0.6719
## Year2012       0.07709    0.12483    0.62    0.5370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.274, Adjusted R-squared:  0.255
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 775 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0141 0.8580 0.9410 0.8940 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      1.17e-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 1.127 1      1.062
## LastAuthorFemale  1.115 1      1.056
## Year              1.221 16      1.006

```

Residuals from first and last author



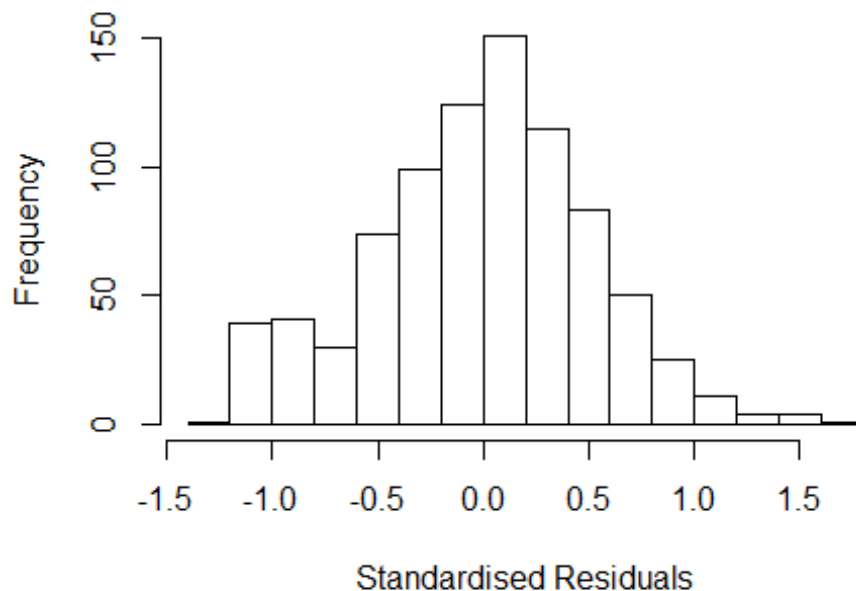
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2408 -0.3279 0.0244 0.3273 1.5763
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89328 0.11454 7.80 1.9e-14 ***
## FirstAuthorFemale1 0.00852 0.03930 0.22 0.8283
## LastAuthorFemale1 -0.13248 0.04524 -2.93 0.0035 **
## Year1997 -0.10975 0.15096 -0.73 0.4674
## Year1998 -0.40910 0.13471 -3.04 0.0025 **
## Year1999 0.11413 0.15213 0.75 0.4533
## Year2000 0.14635 0.14596 1.00 0.3163
## Year2001 -0.04163 0.14495 -0.29 0.7740
## Year2002 0.13006 0.16420 0.79 0.4285
## Year2003 0.09720 0.19452 0.50 0.6174
## Year2004 0.31555 0.13431 2.35 0.0190 *
## Year2005 0.24460 0.13529 1.81 0.0710 .
```

```

## Year2006          0.30695    0.13515    2.27    0.0234 *
## Year2007          0.41664    0.12688    3.28    0.0011 **
## Year2008          0.24632    0.12755    1.93    0.0538 .
## Year2009          0.26537    0.12534    2.12    0.0345 *
## Year2010          0.22652    0.13147    1.72    0.0853 .
## Year2011          0.27539    0.12421    2.22    0.0269 *
## Year2012          0.33896    0.12880    2.63    0.0087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.146, Adjusted R-squared:  0.127
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 778 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.264  0.850   0.952   0.895   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 1.101 1      1.049
## Year              1.101 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
## -1.2009 -0.3221 0.0194 0.3091 1.6100
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8792 0.1172 7.50 1.6e-13 ***
## FirstAuthorFemale1 -0.0136 0.0394 -0.34 0.7305
## Year1997 -0.1255 0.1571 -0.80 0.4248
## Year1998 -0.4066 0.1363 -2.98 0.0029 **
## Year1999 0.1083 0.1545 0.70 0.4834
## Year2000 0.1410 0.1492 0.94 0.3450
## Year2001 -0.0631 0.1490 -0.42 0.6719
## Year2002 0.0916 0.1686 0.54 0.5869
## Year2003 0.0989 0.1931 0.51 0.6088
## Year2004 0.3128 0.1373 2.28 0.0230 *
## Year2005 0.2247 0.1370 1.64 0.1014
## Year2006 0.2983 0.1374 2.17 0.0302 *
```

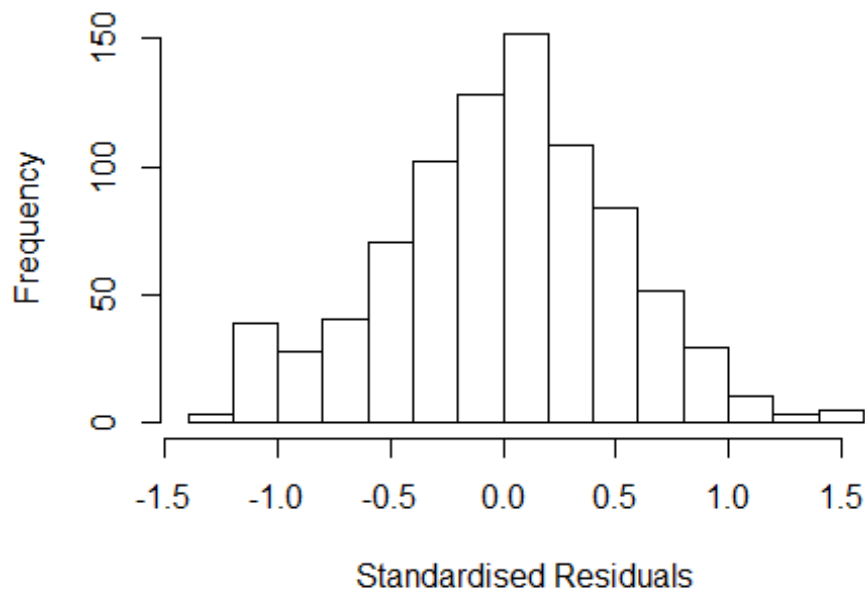


```

## Year2007          0.3926      0.1280      3.07      0.0022 **
## Year2008          0.2454      0.1300      1.89      0.0595 .
## Year2009          0.2519      0.1275      1.98      0.0486 *
## Year2010          0.2271      0.1335      1.70      0.0893 .
## Year2011          0.2592      0.1264      2.05      0.0407 *
## Year2012          0.3217      0.1305      2.46      0.0139 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.476
## Multiple R-squared:  0.137, Adjusted R-squared:  0.12
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 775 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.229  0.849   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      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 1.094 1          1.046
## Year              1.094 16          1.003

```

Residuals from last author



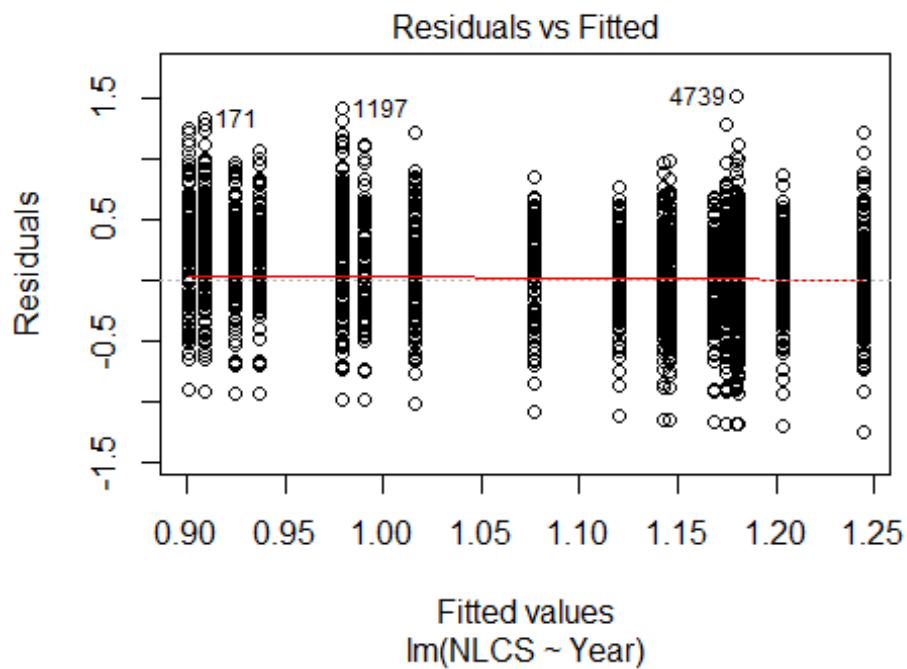
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2357 -0.3296 0.0238 0.3266 1.5858
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8957 0.1142 7.84 1.4e-14 ***
## LastAuthorFemale1 -0.1306 0.0450 -2.90 0.0038 **
## Year1997 -0.1091 0.1530 -0.71 0.4760
## Year1998 -0.4125 0.1347 -3.06 0.0023 **
## Year1999 0.1144 0.1530 0.75 0.4549
## Year2000 0.1473 0.1465 1.01 0.3151
## Year2001 -0.0406 0.1458 -0.28 0.7807
## Year2002 0.1310 0.1646 0.80 0.4263
## Year2003 0.0963 0.1964 0.49 0.6241
## Year2004 0.3162 0.1350 2.34 0.0194 *
## Year2005 0.2453 0.1361 1.80 0.0718 .
## Year2006 0.3087 0.1356 2.28 0.0230 *
```

```

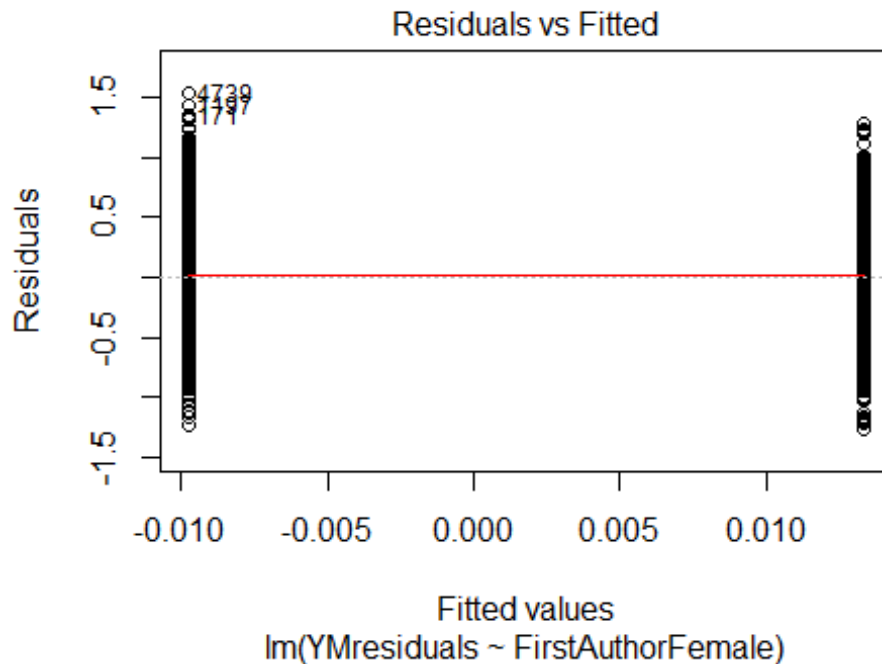
## Year2007          0.4158      0.1271      3.27      0.0011 **
## Year2008          0.2480      0.1284      1.93      0.0537 .
## Year2009          0.2668      0.1259      2.12      0.0344 *
## Year2010          0.2290      0.1321      1.73      0.0834 .
## Year2011          0.2760      0.1247      2.21      0.0272 *
## Year2012          0.3399      0.1295      2.62      0.0088 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.476
## Multiple R-squared:  0.147, Adjusted R-squared:  0.13
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 776 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.244  0.842   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      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: 852"
## [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
## 299 287 233 285 285 244 209 165 216 215 217 198 205 219 293
## 2011 2012
## 273 280
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 195 150 145 186 167 127 146 111 152 163 145 124 140 155 198
## 2011 2012

```

```
## 187 179
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 184 137 130 169 152 117 131 101 140 151 132 110 132 141 178
## 2011 2012
## 170 166
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 170, df = 16, p-value <2e-16
```

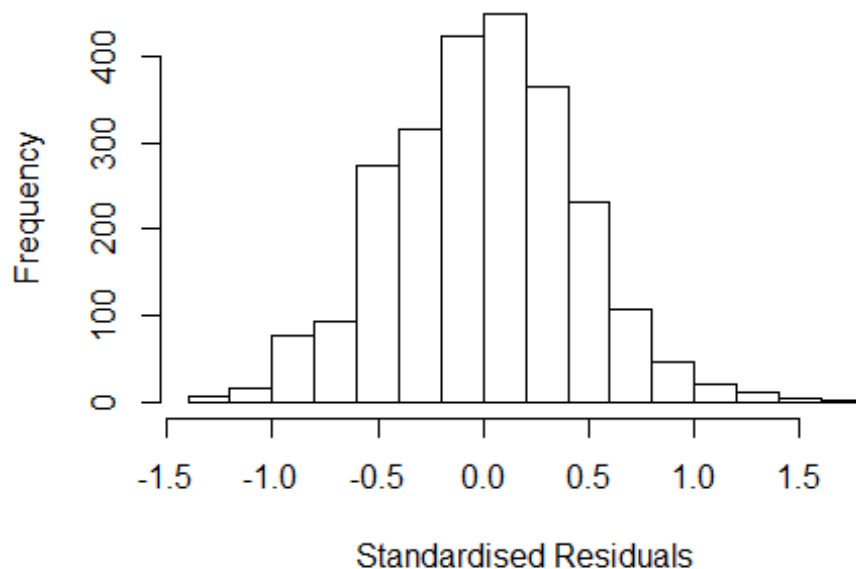


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 5e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 5.24278478346796"
## [1] "Male first author team size 2018 geometric mean: 4.21512614595437"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5600, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.52514148766589"
## [1] "Male last author team size 2018 geometric mean: 4.8596024079597"
##
## 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 1.047 1      1.023
## LastAuthorFemale  1.078 1      1.038
## UniqueAuthors     1.231 4      1.026
## Year              1.297 16     1.008
```

Residuals from first and last author and team size



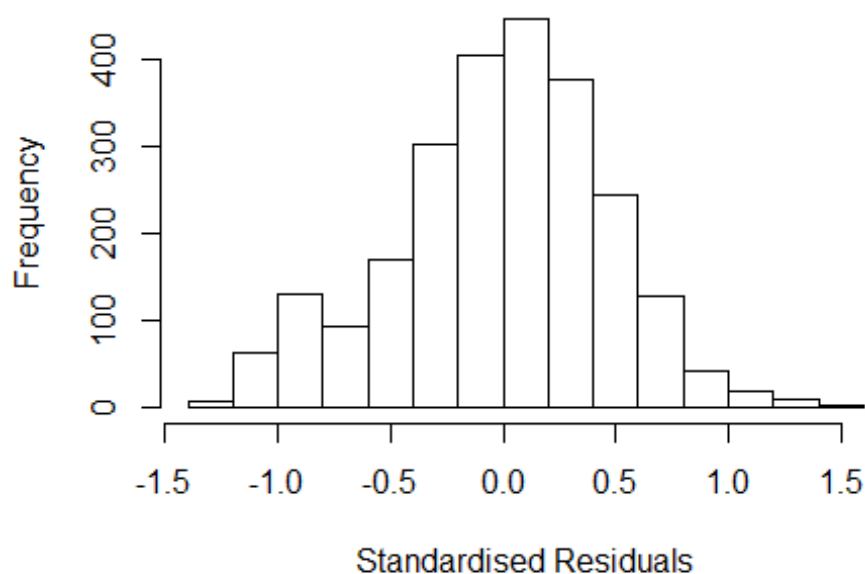
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27140 -0.30379  0.00873  0.29484  1.67838
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.45352    0.05115   8.87 < 2e-16 ***
## FirstAuthorFemale1 -0.00642    0.01829  -0.35  0.72553
## LastAuthorFemale1 -0.00629    0.02160  -0.29  0.77088
## UniqueAuthors2    0.45669    0.04829   9.46 < 2e-16 ***
## UniqueAuthors3    0.58658    0.04546  12.90 < 2e-16 ***
## UniqueAuthors4    0.61084    0.04594  13.30 < 2e-16 ***
## UniqueAuthors5    0.62766    0.04446  14.12 < 2e-16 ***
## Year1997          0.01398    0.06231   0.22  0.82250
## Year1998          0.06359    0.06822   0.93  0.35134
## Year1999          0.05326    0.05482   0.97  0.33141
```

```

## Year2000      0.05270      0.05635      0.94  0.34973
## Year2001      0.06206      0.06030      1.03  0.30351
## Year2002      0.12546      0.05816      2.16  0.03111 *
## Year2003      0.14629      0.05395      2.71  0.00674 **
## Year2004      0.14161      0.04968      2.85  0.00440 **
## Year2005      0.13880      0.04574      3.03  0.00243 **
## Year2006      0.21721      0.04698      4.62  4.0e-06 ***
## Year2007      0.17010      0.05133      3.31  0.00093 ***
## Year2008      0.18075      0.05309      3.40  0.00067 ***
## Year2009      0.15964      0.05072      3.15  0.00167 **
## Year2010      0.19477      0.04727      4.12  3.9e-05 ***
## Year2011      0.19664      0.05242      3.75  0.00018 ***
## Year2012      0.20968      0.04965      4.22  2.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.209, Adjusted R-squared:  0.201
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2230 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0594 0.8630 0.9430 0.8950 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      4.10e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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      1.020
## LastAuthorFemale 1.046 1      1.023
## 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
## -1.2545 -0.2894  0.0179  0.3084  1.5054
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9161    0.0500   18.33 < 2e-16 ***
## FirstAuthorFemale1  0.0236    0.0195    1.21  0.22523
## LastAuthorFemale1 -0.0421    0.0236   -1.79  0.07417 .
## Year1997         0.0199    0.0793    0.25  0.80177
## Year1998         0.0968    0.0746    1.30  0.19442
## Year1999         0.0867    0.0697    1.24  0.21401
## Year2000         0.0555    0.0682    0.81  0.41537
## Year2001         0.0822    0.0792    1.04  0.29946
## Year2002         0.1539    0.0684    2.25  0.02447 *
## Year2003         0.1964    0.0642    3.06  0.00223 **
## Year2004         0.2231    0.0598    3.73  0.00019 ***
## Year2005         0.2165    0.0560    3.87  0.00011 ***
```

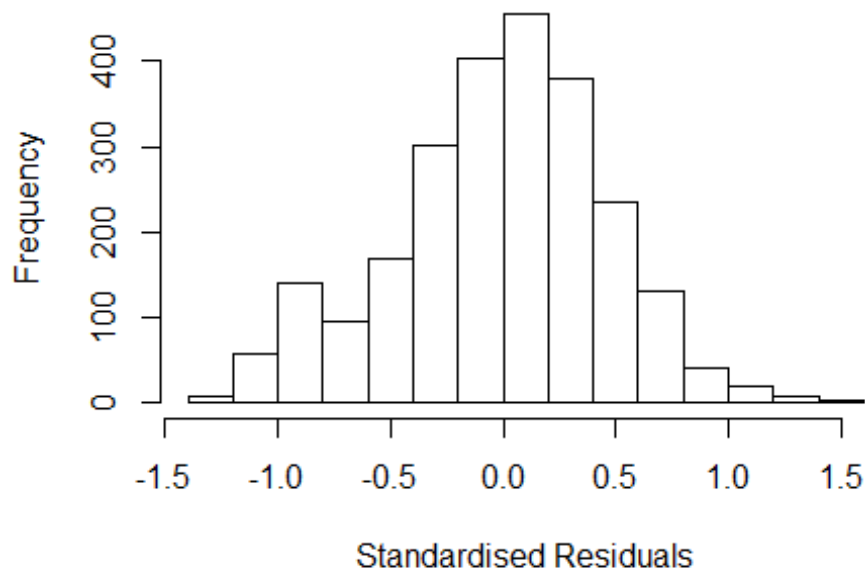


```

## Year2006          0.3029      0.0575      5.27  1.5e-07 ***
## Year2007          0.2650      0.0616      4.30  1.7e-05 ***
## Year2008          0.2569      0.0592      4.34  1.5e-05 ***
## Year2009          0.2445      0.0605      4.04  5.4e-05 ***
## Year2010          0.2883      0.0574      5.02  5.5e-07 ***
## Year2011          0.2805      0.0629      4.46  8.5e-06 ***
## Year2012          0.3148      0.0594      5.30  1.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0496, Adjusted R-squared:  0.0425
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 221 weights are ~= 1. The remaining 2220 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.213  0.854  0.948  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      4.10e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



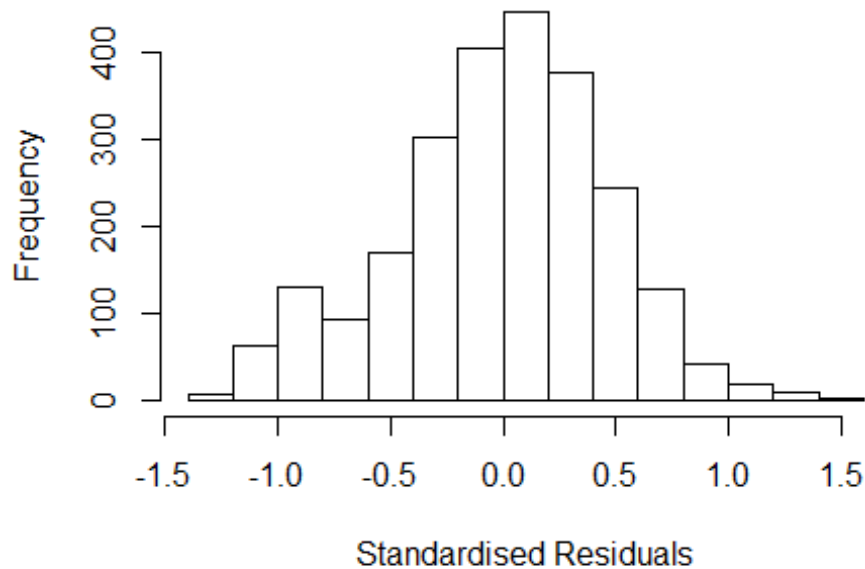
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2429 -0.2891 0.0172 0.3001 1.5188
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9095 0.0497 18.29 < 2e-16 ***
## FirstAuthorFemale1 0.0213 0.0197 1.08 0.27992
## Year1997 0.0189 0.0793 0.24 0.81192
## Year1998 0.0934 0.0744 1.25 0.20962
## Year1999 0.0870 0.0695 1.25 0.21029
## Year2000 0.0550 0.0682 0.81 0.41957
## Year2001 0.0793 0.0790 1.00 0.31573
## Year2002 0.1501 0.0684 2.20 0.02824 *
## Year2003 0.1913 0.0638 3.00 0.00274 **
## Year2004 0.2228 0.0596 3.74 0.00019 ***
## Year2005 0.2150 0.0556 3.86 0.00011 ***
## Year2006 0.3023 0.0572 5.28 1.4e-07 ***
```

```

## Year2007          0.2634      0.0612      4.30  1.8e-05 ***
## Year2008          0.2553      0.0590      4.32  1.6e-05 ***
## Year2009          0.2394      0.0601      3.99  6.9e-05 ***
## Year2010          0.2843      0.0570      4.99  6.6e-07 ***
## Year2011          0.2736      0.0624      4.38  1.2e-05 ***
## Year2012          0.3121      0.0591      5.28  1.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0482, Adjusted R-squared:  0.0416
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 212 weights are ~= 1. The remaining 2229 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.855  0.948  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      4.10e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



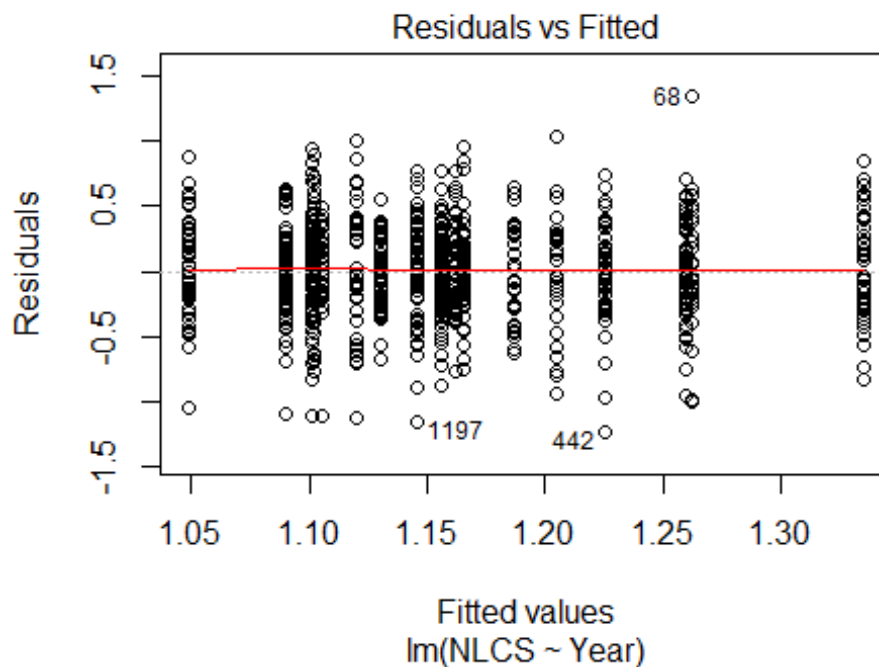
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2417 -0.2875 0.0184 0.3091 1.4948
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9239 0.0496 18.62 < 2e-16 ***
## LastAuthorFemale1 -0.0403 0.0237 -1.70 0.08948 .
## Year1997 0.0219 0.0794 0.28 0.78293
## Year1998 0.0986 0.0746 1.32 0.18657
## Year1999 0.0868 0.0699 1.24 0.21419
## Year2000 0.0564 0.0681 0.83 0.40807
## Year2001 0.0848 0.0792 1.07 0.28485
## Year2002 0.1554 0.0684 2.27 0.02321 *
## Year2003 0.2001 0.0641 3.12 0.00184 **
## Year2004 0.2240 0.0600 3.74 0.00019 ***
## Year2005 0.2191 0.0561 3.90 9.7e-05 ***
## Year2006 0.3059 0.0576 5.31 1.2e-07 ***
```

```

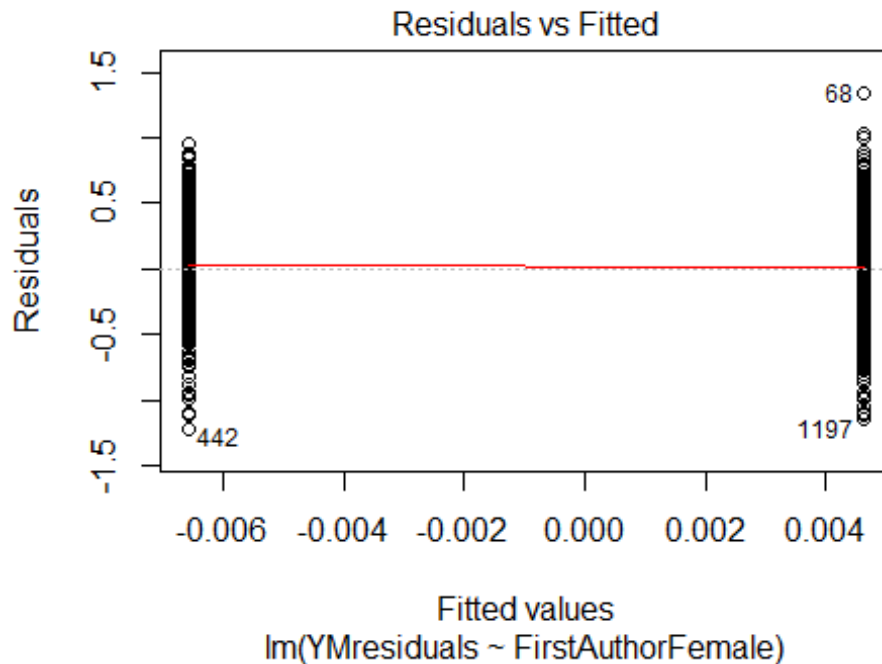
## Year2007          0.2665      0.0617      4.32  1.6e-05 ***
## Year2008          0.2593      0.0594      4.37  1.3e-05 ***
## Year2009          0.2485      0.0606      4.10  4.3e-05 ***
## Year2010          0.2901      0.0576      5.04  5.1e-07 ***
## Year2011          0.2832      0.0631      4.49  7.4e-06 ***
## Year2012          0.3178      0.0596      5.33  1.0e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0422
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2221 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.219  0.854  0.948  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      4.10e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2441"
## [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
##   64   61   58   65   70   90   76   83   59   83  114   83   89   95   98
## 2011 2012
##   96   94
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   38   34   45   36   45   44   53   38   51   75   58   69   71   75
## 2011 2012

```

```
## 70 63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 28 34 31 41 33 45 41 49 35 45 69 53 67 63 65
## 2011 2012
## 66 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 = 35, df = 16, p-value = 0.004
```



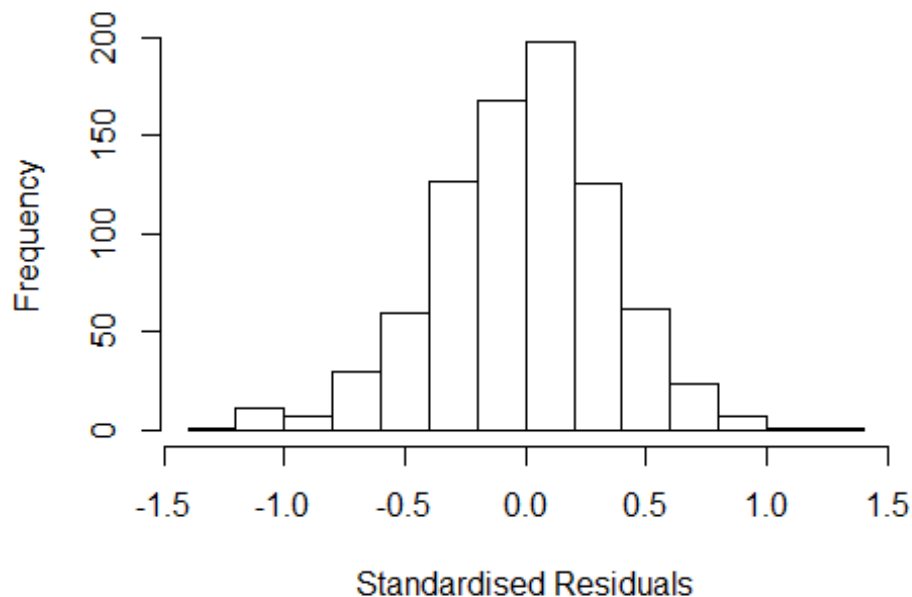
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00032, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 4.8710345857669"
## [1] "Male first author team size 2018 geometric mean: 3.7689057315738"
## 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 = 1200, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.92263366809729"
## [1] "Male last author team size 2018 geometric mean: 4.58970385313413"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 690, 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	1.129	1	1.063
LastAuthorFemale	1.079	1	1.039
UniqueAuthors	1.633	4	1.063
Year	1.786	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
## -1.283 -0.227 0.011 0.217 1.315
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13325 0.10106 11.21 <2e-16 ***
## FirstAuthorFemale1 -0.03371 0.02656 -1.27 0.205
## LastAuthorFemale1 0.04290 0.02941 1.46 0.145
## UniqueAuthors2 0.04082 0.07182 0.57 0.570
## UniqueAuthors3 0.15461 0.07086 2.18 0.029 *
## UniqueAuthors4 0.13445 0.07080 1.90 0.058 .
## UniqueAuthors5 0.16792 0.06987 2.40 0.016 *
## Year1997 -0.06367 0.11976 -0.53 0.595
## Year1998 0.00870 0.11610 0.07 0.940
## Year1999 0.06322 0.10335 0.61 0.541
```

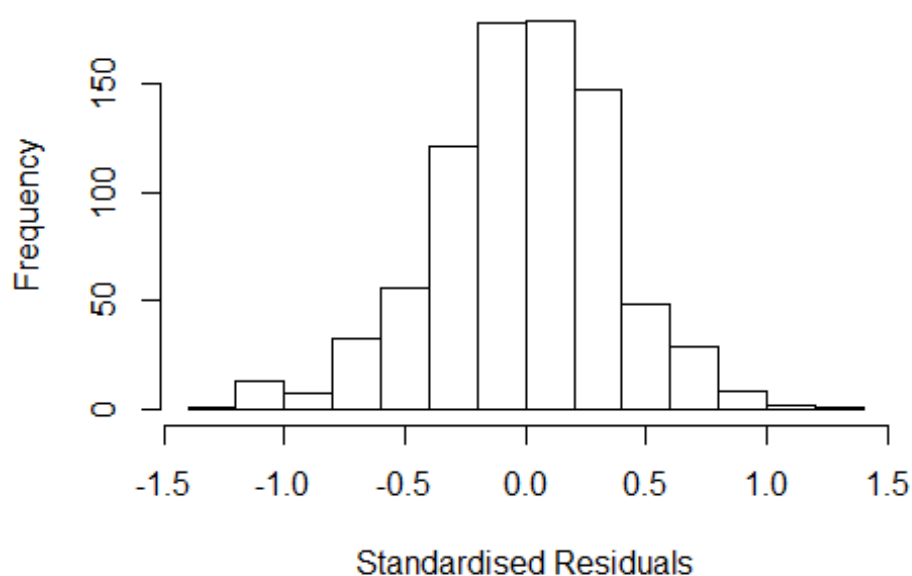


```

## Year2000      -0.05077    0.10454   -0.49    0.627
## Year2001      0.02838    0.09813    0.29    0.772
## Year2002     -0.15558    0.10146   -1.53    0.126
## Year2003     -0.09213    0.09195   -1.00    0.317
## Year2004     -0.13109    0.08933   -1.47    0.143
## Year2005     -0.10490    0.09135   -1.15    0.251
## Year2006     -0.14415    0.09016   -1.60    0.110
## Year2007     -0.11966    0.09305   -1.29    0.199
## Year2008     -0.06444    0.09437   -0.68    0.495
## Year2009     -0.10397    0.09407   -1.11    0.269
## Year2010     -0.08086    0.09303   -0.87    0.385
## Year2011     -0.16078    0.09359   -1.72    0.086 .
## Year2012      0.00594    0.09304    0.06    0.949
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared:  0.0606, Adjusted R-squared:  0.0347
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 763 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.108  0.871  0.955  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      1.22e-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 1.094 1      1.046
## LastAuthorFemale  1.080 1      1.039
## Year              1.179 16      1.005

```

Residuals from first and last author



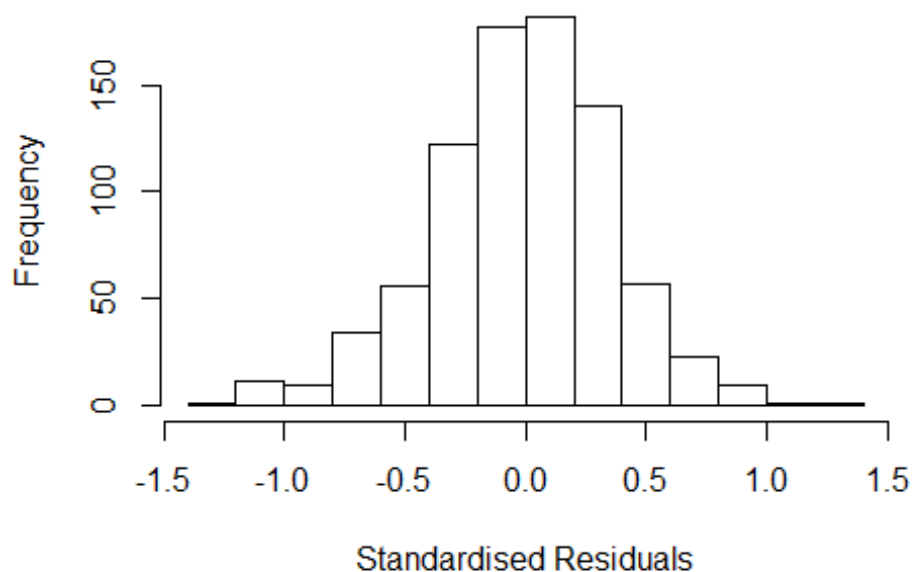
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2571 -0.2279 0.0047 0.2252 1.3602
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2428 0.0849 14.63 <2e-16 ***
## FirstAuthorFemale1 -0.0189 0.0263 -0.72 0.473
## LastAuthorFemale1 0.0508 0.0297 1.71 0.087 .
## Year1997 -0.0926 0.1257 -0.74 0.462
## Year1998 -0.0139 0.1221 -0.11 0.910
## Year1999 0.0647 0.1080 0.60 0.549
## Year2000 -0.0695 0.1081 -0.64 0.520
## Year2001 0.0333 0.1038 0.32 0.749
## Year2002 -0.1552 0.1047 -1.48 0.139
## Year2003 -0.0933 0.0984 -0.95 0.343
## Year2004 -0.1330 0.0939 -1.42 0.157
## Year2005 -0.1022 0.0962 -1.06 0.289
```

```

## Year2006          -0.1306      0.0954   -1.37    0.171
## Year2007          -0.1107      0.0966   -1.15    0.252
## Year2008          -0.0690      0.0981   -0.70    0.482
## Year2009          -0.1088      0.0983   -1.11    0.269
## Year2010          -0.0698      0.0974   -0.72    0.474
## Year2011          -0.1544      0.0988   -1.56    0.119
## Year2012           0.0204      0.0988    0.21    0.836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared:  0.0351, Adjusted R-squared:  0.0135
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 755 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.090  0.878  0.954  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      1.22e-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 1.097 1      1.047
## Year              1.097 16      1.003

```

Residuals from first author



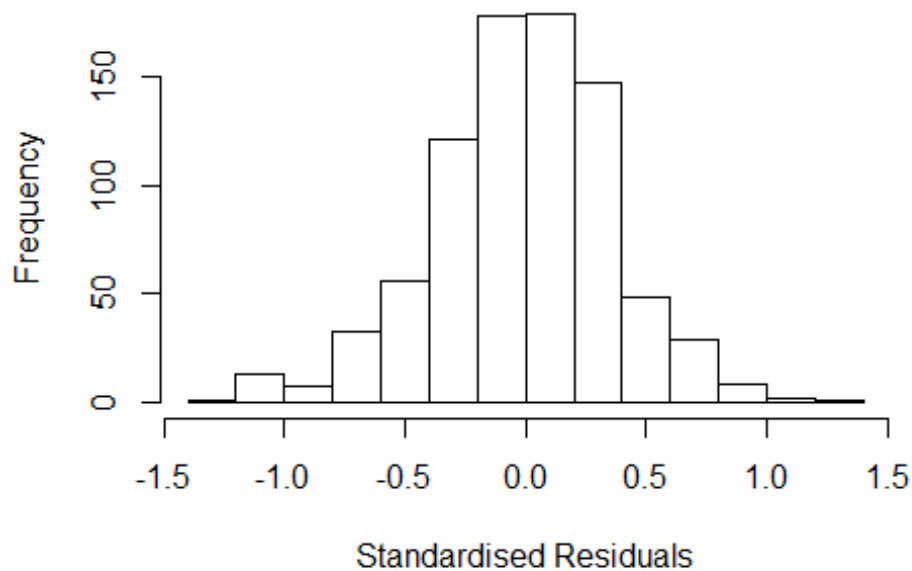
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26781 -0.23375 0.00288 0.23037 1.34714
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2559 0.0861 14.58 <2e-16 ***
## FirstAuthorFemale1 -0.0155 0.0265 -0.58 0.56
## Year1997 -0.0884 0.1258 -0.70 0.48
## Year1998 -0.0203 0.1214 -0.17 0.87
## Year1999 0.0632 0.1085 0.58 0.56
## Year2000 -0.0745 0.1095 -0.68 0.50
## Year2001 0.0274 0.1041 0.26 0.79
## Year2002 -0.1635 0.1062 -1.54 0.12
## Year2003 -0.0978 0.0990 -0.99 0.32
## Year2004 -0.1379 0.0943 -1.46 0.14
## Year2005 -0.1063 0.0971 -1.10 0.27
## Year2006 -0.1278 0.0958 -1.33 0.18
```

```

## Year2007          -0.1082      0.0966   -1.12      0.26
## Year2008          -0.0680      0.0984   -0.69      0.49
## Year2009          -0.1128      0.0993   -1.14      0.26
## Year2010          -0.0670      0.0976   -0.69      0.49
## Year2011          -0.1535      0.0996   -1.54      0.12
## Year2012           0.0202      0.0992    0.20      0.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.346
## Multiple R-squared:  0.0314, Adjusted R-squared:  0.011
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 57 weights are ~= 1. The remaining 766 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0969 0.8750 0.9540 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      1.22e-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 1.08 1      1.039
## Year              1.08 16      1.002

```

Residuals from last author



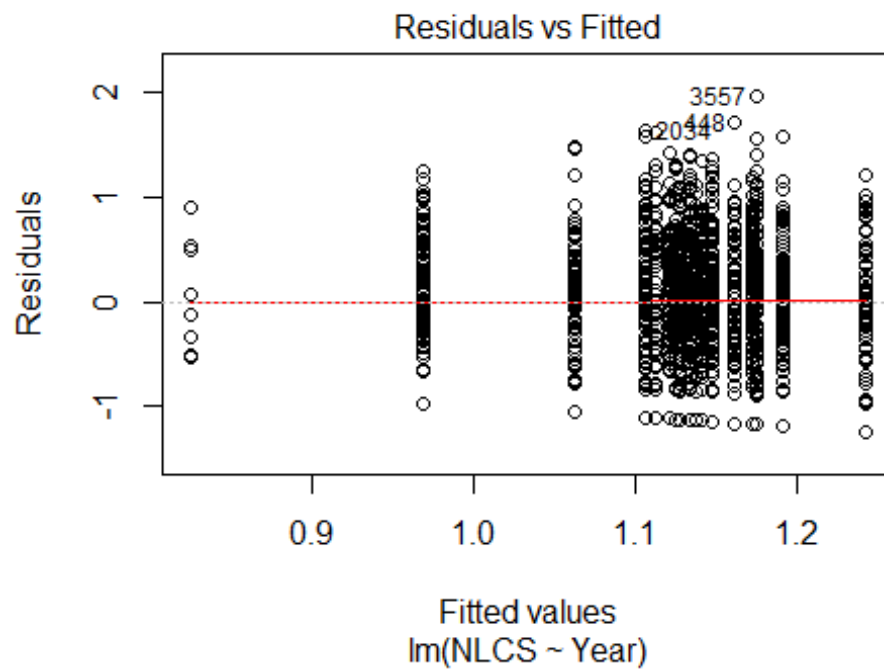
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2672 -0.2231 0.0042 0.2250 1.3612
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2418 0.0851 14.60 <2e-16 ***
## LastAuthorFemale1 0.0492 0.0299 1.65 0.100 .
## Year1997 -0.0983 0.1255 -0.78 0.434
## Year1998 -0.0182 0.1216 -0.15 0.881
## Year1999 0.0615 0.1081 0.57 0.570
## Year2000 -0.0752 0.1081 -0.70 0.487
## Year2001 0.0254 0.1033 0.25 0.806
## Year2002 -0.1608 0.1046 -1.54 0.125
## Year2003 -0.0998 0.0984 -1.01 0.311
## Year2004 -0.1418 0.0935 -1.52 0.130
## Year2005 -0.1130 0.0954 -1.19 0.236
## Year2006 -0.1362 0.0949 -1.43 0.152
```

```

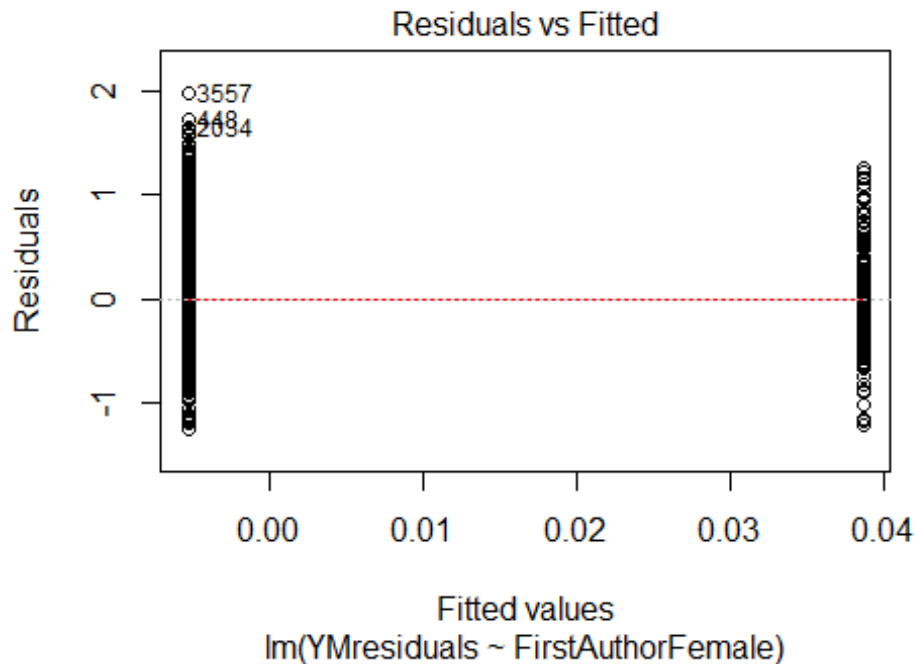
## Year2007          -0.1133      0.0967    -1.17     0.241
## Year2008          -0.0760      0.0976    -0.78     0.436
## Year2009          -0.1155      0.0981    -1.18     0.239
## Year2010          -0.0772      0.0969    -0.80     0.426
## Year2011          -0.1623      0.0984    -1.65     0.099 .
## Year2012           0.0124      0.0986     0.13     0.900
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0142
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 750 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0906 0.8770 0.9530 0.8950 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      1.22e-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: 823"
## [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
##   26  253  251  267  169  253  202  274  234  239  278  265  280  281  270
## 2011 2012
##   285  315
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9   83   83   91   54   63   72  103   82   95  106  103  132  118  135
## 2011 2012

```

```
## 140 167
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 69 69 85 52 53 61 82 65 74 87 86 113 101 120
## 2011 2012
## 124 147
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```

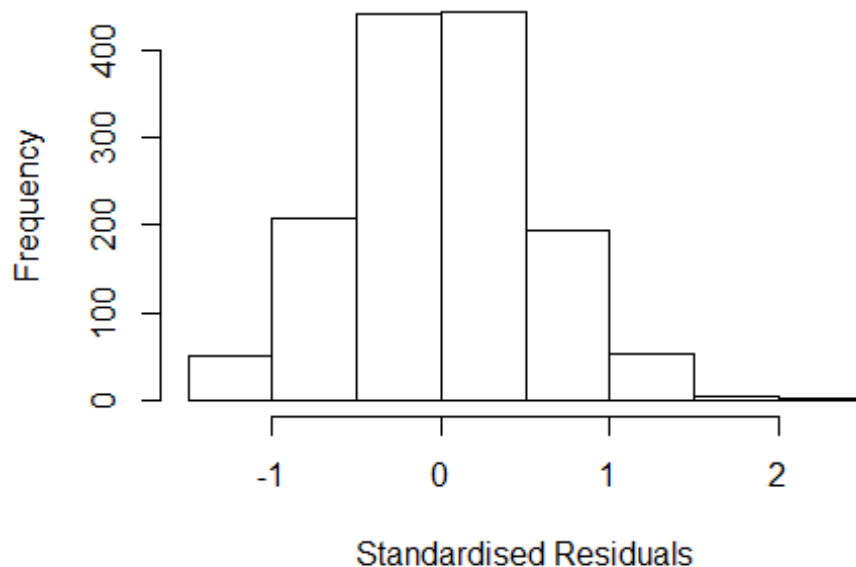


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```

```
## [1] "Female first author team size 2018 geometric mean: 4.85920895904566"
## [1] "Male first author team size 2018 geometric mean: 2.59350450639016"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3000, p-value = 5e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.38496690797842"
## [1] "Male last author team size 2018 geometric mean: 2.93969539568311"
##
## 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] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.127 1          1.061
## LastAuthorFemale  1.076 1          1.038
## UniqueAuthors    1.295 4          1.033
## Year              1.329 16         1.009
```

Residuals from first and last author and team size



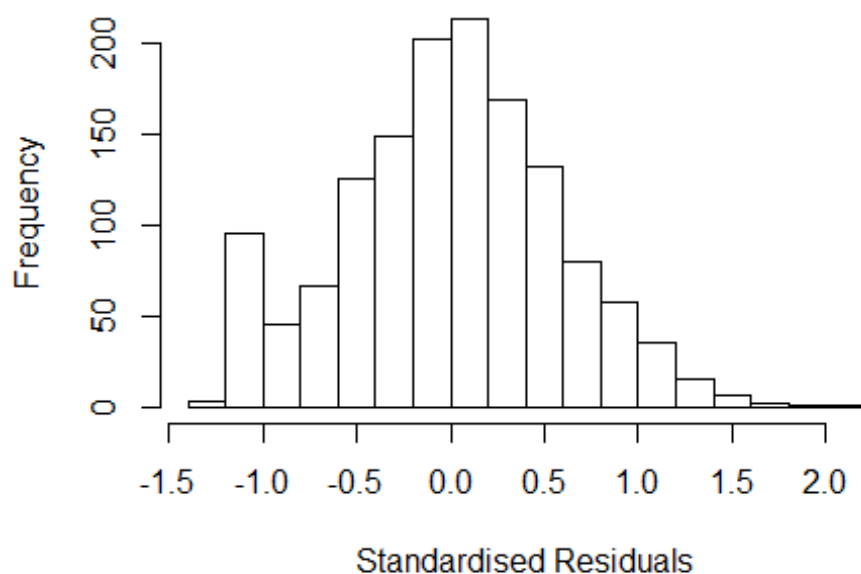
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42759 -0.33369 -0.00203 0.35846 2.00265
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6464 0.1488 4.35 1.5e-05 ***
## FirstAuthorFemale1 -0.0343 0.0451 -0.76 0.447
## LastAuthorFemale1 0.0291 0.0570 0.51 0.610
## UniqueAuthors2 0.2221 0.0420 5.28 1.5e-07 ***
## UniqueAuthors3 0.3383 0.0435 7.77 1.5e-14 ***
## UniqueAuthors4 0.3523 0.0534 6.60 5.9e-11 ***
## UniqueAuthors5 0.4832 0.0585 8.26 3.4e-16 ***
## Year1997 0.2777 0.1658 1.67 0.094 .
## Year1998 0.2413 0.1637 1.47 0.141
## Year1999 0.3335 0.1582 2.11 0.035 *
```

```

## Year2000          0.2775      0.1648      1.68      0.092 .
## Year2001          0.2942      0.1960      1.50      0.134
## Year2002          0.3680      0.1669      2.20      0.028 *
## Year2003          0.2219      0.1698      1.31      0.192
## Year2004          0.2164      0.1658      1.31      0.192
## Year2005          0.2066      0.1618      1.28      0.202
## Year2006          0.2577      0.1607      1.60      0.109
## Year2007          0.2699      0.1638      1.65      0.100 .
## Year2008          0.1339      0.1582      0.85      0.397
## Year2009          0.2900      0.1618      1.79      0.073 .
## Year2010          0.2767      0.1601      1.73      0.084 .
## Year2011          0.2864      0.1577      1.82      0.070 .
## Year2012          0.2980      0.1564      1.91      0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0875, Adjusted R-squared:  0.0729
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1275 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.125  0.858  0.952   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      7.16e-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 1.082 1      1.040
## LastAuthorFemale  1.039 1      1.019
## Year              1.091 16      1.003

```

Residuals from first and last author



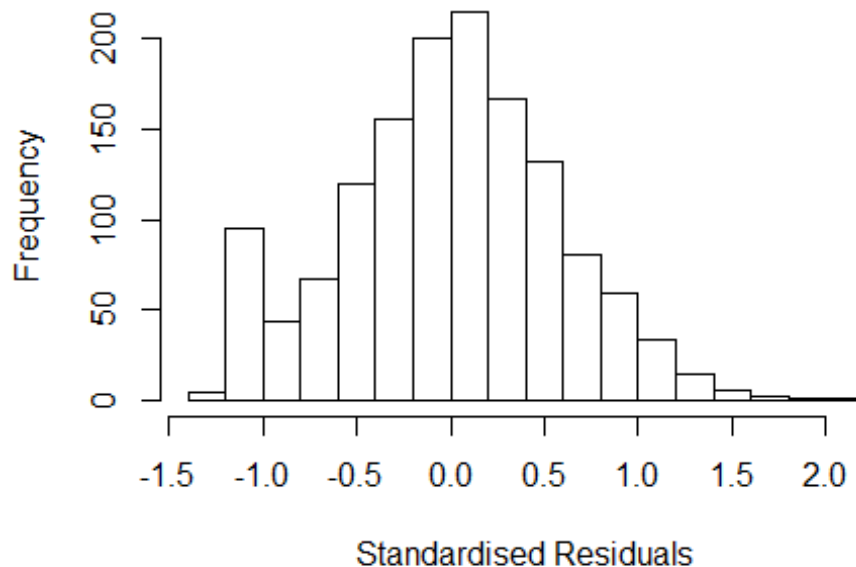
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2008 -0.3754 0.0112 0.3787 2.0308
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7260 0.1883 3.86 0.00012 ***
## FirstAuthorFemale1 0.0225 0.0467 0.48 0.62961
## LastAuthorFemale1 0.0631 0.0613 1.03 0.30276
## Year1997 0.3784 0.2014 1.88 0.06041 .
## Year1998 0.3271 0.2012 1.63 0.10415
## Year1999 0.4129 0.1971 2.10 0.03632 *
## Year2000 0.3945 0.2036 1.94 0.05291 .
## Year2001 0.3894 0.2236 1.74 0.08179 .
## Year2002 0.4748 0.2029 2.34 0.01941 *
## Year2003 0.3526 0.2060 1.71 0.08720 .
## Year2004 0.3095 0.2019 1.53 0.12563
## Year2005 0.3186 0.1980 1.61 0.10782
```

```

## Year2006          0.3728      0.1988      1.88  0.06099 .
## Year2007          0.3842      0.2010      1.91  0.05613 .
## Year2008          0.2495      0.1959      1.27  0.20302
## Year2009          0.4194      0.1991      2.11  0.03530 *
## Year2010          0.4124      0.1975      2.09  0.03697 *
## Year2011          0.4652      0.1941      2.40  0.01670 *
## Year2012          0.4495      0.1939      2.32  0.02057 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0153, Adjusted R-squared:  0.00242
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1260 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.158  0.859  0.949  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      7.16e-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 1.064 1          1.031
## Year              1.064 16          1.002

```

Residuals from first author



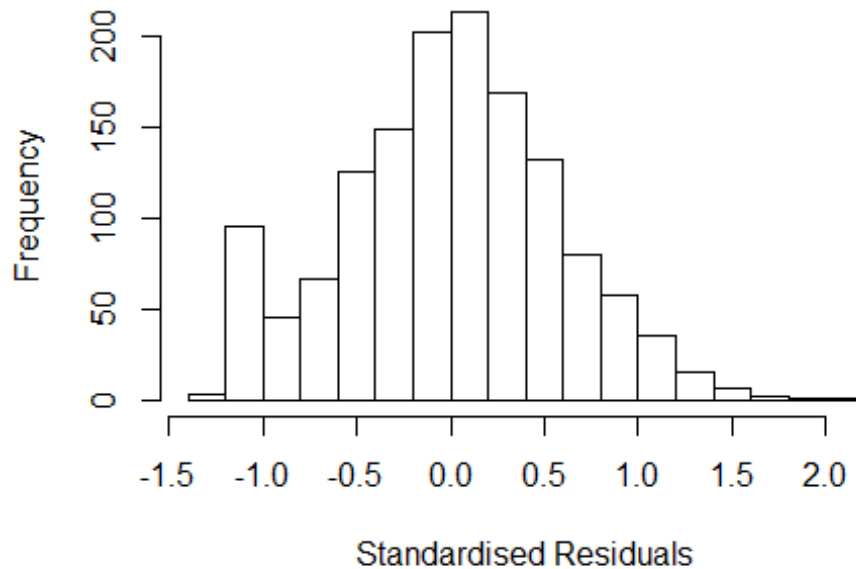
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2127 -0.3694 0.0097 0.3772 2.0285
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7260 0.1883 3.86 0.00012 ***
## FirstAuthorFemale1 0.0335 0.0464 0.72 0.47015
## Year1997 0.3838 0.2014 1.91 0.05688 .
## Year1998 0.3299 0.2010 1.64 0.10097
## Year1999 0.4158 0.1970 2.11 0.03501 *
## Year2000 0.3987 0.2038 1.96 0.05055 .
## Year2001 0.3908 0.2238 1.75 0.08101 .
## Year2002 0.4813 0.2024 2.38 0.01755 *
## Year2003 0.3533 0.2061 1.71 0.08679 .
## Year2004 0.3119 0.2019 1.54 0.12267
## Year2005 0.3246 0.1979 1.64 0.10119
## Year2006 0.3769 0.1988 1.90 0.05826 .
```

```

## Year2007          0.3864      0.2010      1.92  0.05478 .
## Year2008          0.2531      0.1959      1.29  0.19653
## Year2009          0.4232      0.1990      2.13  0.03366 *
## Year2010          0.4168      0.1975      2.11  0.03505 *
## Year2011          0.4697      0.1940      2.42  0.01562 *
## Year2012          0.4532      0.1939      2.34  0.01958 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00232
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1259 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.159  0.859  0.949  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      7.16e-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 1.022 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
## -1.20188 -0.37334 0.00932 0.37633 2.02815
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7259 0.1883 3.85 0.00012 ***
## LastAuthorFemale1 0.0687 0.0607 1.13 0.25782
## Year1997 0.3800 0.2013 1.89 0.05925 .
## Year1998 0.3292 0.2012 1.64 0.10209
## Year1999 0.4155 0.1969 2.11 0.03505 *
## Year2000 0.3974 0.2036 1.95 0.05112 .
## Year2001 0.3917 0.2234 1.75 0.07973 .
## Year2002 0.4759 0.2029 2.35 0.01912 *
## Year2003 0.3554 0.2058 1.73 0.08451 .
## Year2004 0.3119 0.2018 1.55 0.12233
## Year2005 0.3200 0.1980 1.62 0.10620
## Year2006 0.3740 0.1988 1.88 0.06021 .
```

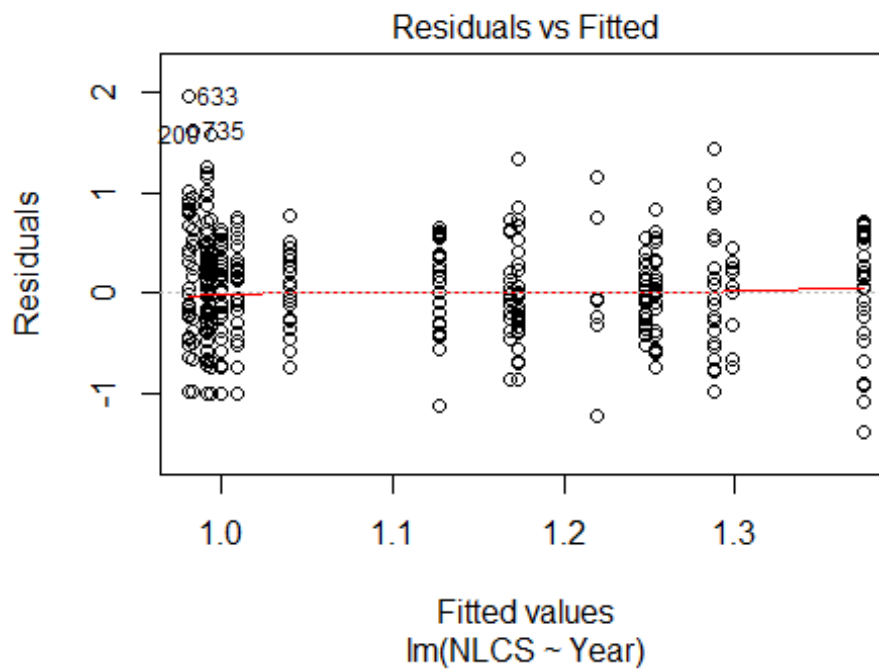


```

## Year2007          0.3869      0.2008      1.93  0.05422 .
## Year2008          0.2520      0.1960      1.29  0.19867
## Year2009          0.4216      0.1991      2.12  0.03438 *
## Year2010          0.4163      0.1972      2.11  0.03499 *
## Year2011          0.4683      0.1939      2.41  0.01587 *
## Year2012          0.4514      0.1938      2.33  0.02002 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.00303
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1260 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.159  0.860  0.950  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      7.16e-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: 1396"
## [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
##   89   71   75  109   96   95   86   93   92   90  108   95   84   72   49
## 2011 2012
##   47   72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   14   18   24   37   30   19   26   25   28   22   26   22   32   11
## 2011 2012

```

```
## 22 28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 12 14 19 34 26 13 21 19 19 17 19 16 21 5
## 2011 2012
## 19 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 = 49, df = 16, p-value = 3e-05
```



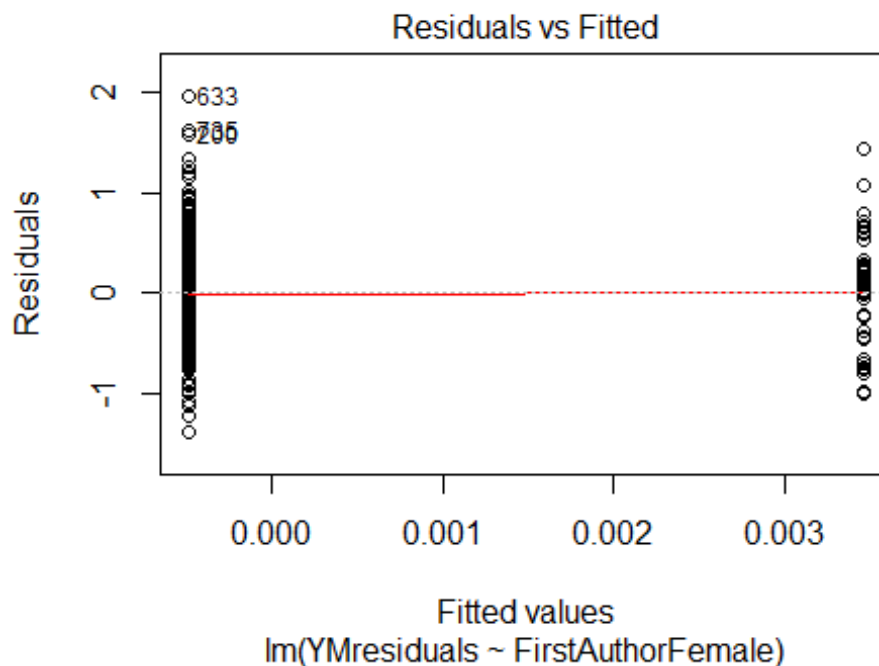
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 2.85173947486242"
## [1] "Male first author team size 2018 geometric mean: 2.41386134202753"

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

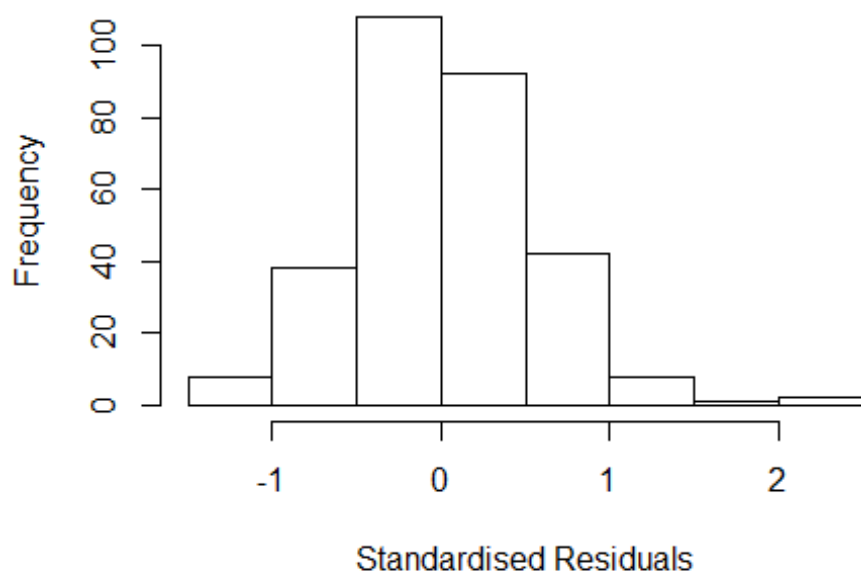
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 73, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.73205080756888"
## [1] "Male last author team size 2018 geometric mean: 2.62335744811587"

## 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 = 15, 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 1.247  1      1.117
## LastAuthorFemale  1.286  1      1.134
## UniqueAuthors    2.694  4      1.132
## Year              3.347 16      1.038
```

Residuals from first and last author and team size



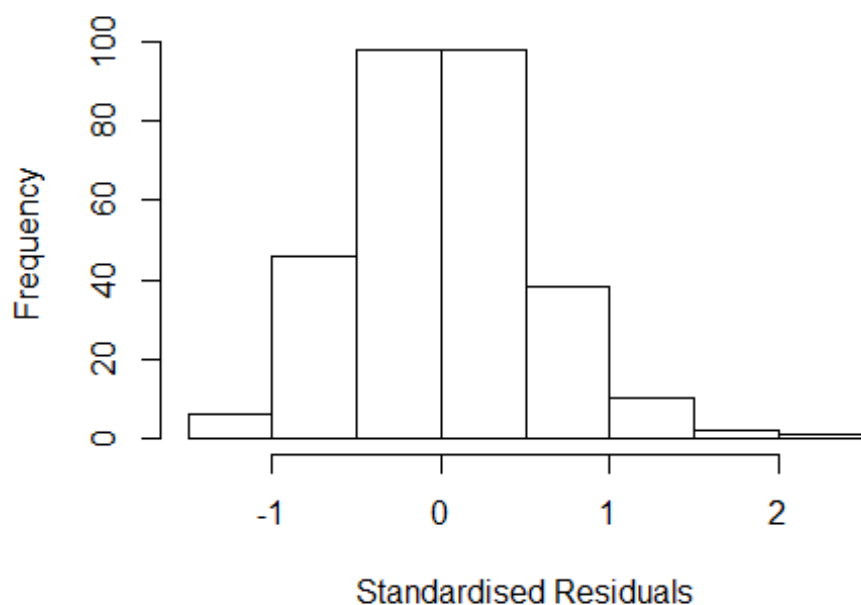
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3839 -0.3169 -0.0167 0.3442 2.1584
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2194 0.6200 1.97 0.0502 .
## FirstAuthorFemale1 0.0432 0.1083 0.40 0.6902
## LastAuthorFemale1 0.0661 0.0971 0.68 0.4961
## UniqueAuthors2 0.1229 0.1002 1.23 0.2212
## UniqueAuthors3 0.1645 0.0900 1.83 0.0689 .
## UniqueAuthors4 0.2700 0.1003 2.69 0.0075 **
## UniqueAuthors5 0.3262 0.1117 2.92 0.0038 **
## Year1997 -0.0888 0.6243 -0.14 0.8871
## Year1998 -0.4625 0.6241 -0.74 0.4593
## Year1999 -0.1342 0.6265 -0.21 0.8305
```

```

## Year2000          -0.3559      0.6179    -0.58    0.5651
## Year2001          -0.4420      0.6304    -0.70    0.4838
## Year2002          -0.7748      0.6358    -1.22    0.2240
## Year2003           0.0390      0.6221     0.06    0.9500
## Year2004          -0.2897      0.6168    -0.47    0.6389
## Year2005          -0.4366      0.6109    -0.71    0.4755
## Year2006          -0.0777      0.6120    -0.13    0.8990
## Year2007          -0.4797      0.6104    -0.79    0.4326
## Year2008          -0.1361      0.6080    -0.22    0.8230
## Year2009          -0.3789      0.6138    -0.62    0.5375
## Year2010          -0.1683      0.6336    -0.27    0.7907
## Year2011          -0.4067      0.6116    -0.66    0.5067
## Year2012          -0.3692      0.6190    -0.60    0.5513
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.148, Adjusted R-squared:  0.0795
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 264 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0254 0.8670 0.9460 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          3.34e-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 1.170 1 1.082
## LastAuthorFemale 1.179 1 1.086
## Year 1.358 16 1.010

```

Residuals from first and last author



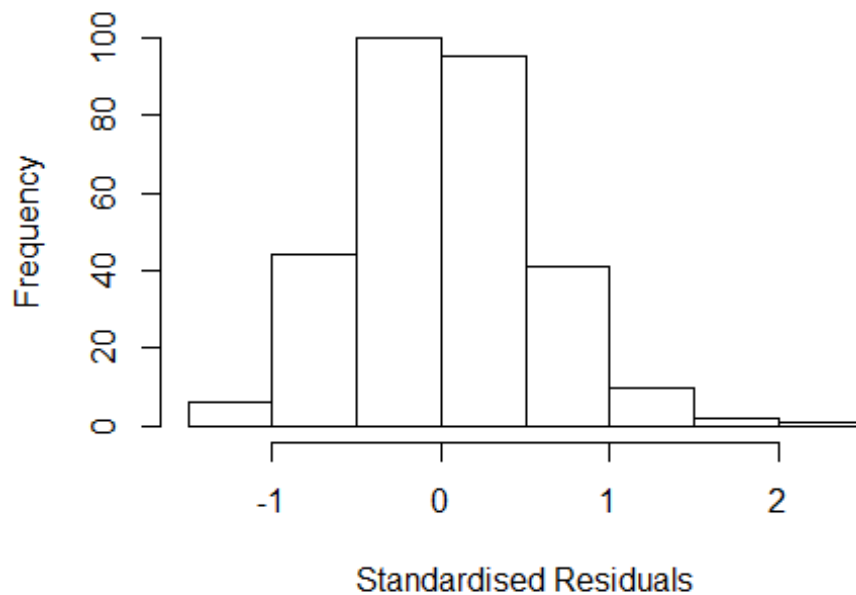
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41229 -0.32324 -0.00952  0.37103  2.03348
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3490     0.5088   2.65  0.0085 **
## FirstAuthorFemale1  0.0371     0.1075   0.35  0.7301
## LastAuthorFemale1  0.0602     0.0949   0.63  0.5262
## Year1997          -0.1638     0.5232  -0.31  0.7545
## Year1998          -0.4929     0.5296  -0.93  0.3528
## Year1999          -0.1161     0.5336  -0.22  0.8280
## Year2000          -0.3634     0.5212  -0.70  0.4862
## Year2001          -0.4494     0.5410  -0.83  0.4068
## Year2002          -0.6959     0.5598  -1.24  0.2148
## Year2003           0.0633     0.5278   0.12  0.9046
## Year2004          -0.2813     0.5226  -0.54  0.5908
## Year2005          -0.4127     0.5188  -0.80  0.4270
```

```

## Year2006          -0.0371      0.5185   -0.07   0.9430
## Year2007          -0.4452      0.5120   -0.87   0.3853
## Year2008          -0.0814      0.5133   -0.16   0.8742
## Year2009          -0.3809      0.5161   -0.74   0.4611
## Year2010          -0.1287      0.5359   -0.24   0.8103
## Year2011          -0.3454      0.5141   -0.67   0.5022
## Year2012          -0.2920      0.5212   -0.56   0.5758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.0524
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 277 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0932 0.8750 0.9510 0.9010 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      3.34e-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 1.165 1      1.079
## Year              1.165 16      1.005

```

Residuals from first author



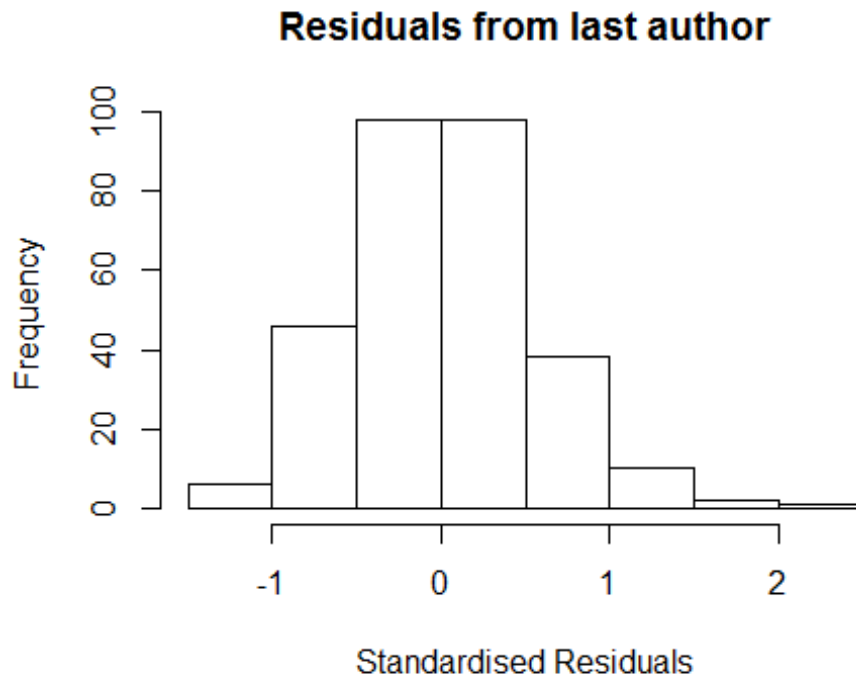
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41375 -0.33309 -0.00109 0.36765 2.02486
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3648 0.5223 2.61 0.0095 **
## FirstAuthorFemale1 0.0491 0.1074 0.46 0.6479
## Year1997 -0.1699 0.5384 -0.32 0.7525
## Year1998 -0.4966 0.5450 -0.91 0.3629
## Year1999 -0.1311 0.5458 -0.24 0.8103
## Year2000 -0.3696 0.5358 -0.69 0.4909
## Year2001 -0.4567 0.5566 -0.82 0.4127
## Year2002 -0.6979 0.5726 -1.22 0.2239
## Year2003 0.0489 0.5415 0.09 0.9281
## Year2004 -0.2920 0.5366 -0.54 0.5868
## Year2005 -0.4303 0.5317 -0.81 0.4190
## Year2006 -0.0510 0.5322 -0.10 0.9237
```



```

## Year2007          -0.4456      0.5281   -0.84   0.3995
## Year2008          -0.0938      0.5274   -0.18   0.8589
## Year2009          -0.3948      0.5295   -0.75   0.4566
## Year2010          -0.1448      0.5486   -0.26   0.7920
## Year2011          -0.3538      0.5287   -0.67   0.5040
## Year2012          -0.2962      0.5362   -0.55   0.5811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.107, Adjusted R-squared:  0.053
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 274 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.877  0.952  0.901  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      3.34e-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 1.177 1          1.085
## Year            1.177 16          1.005

```



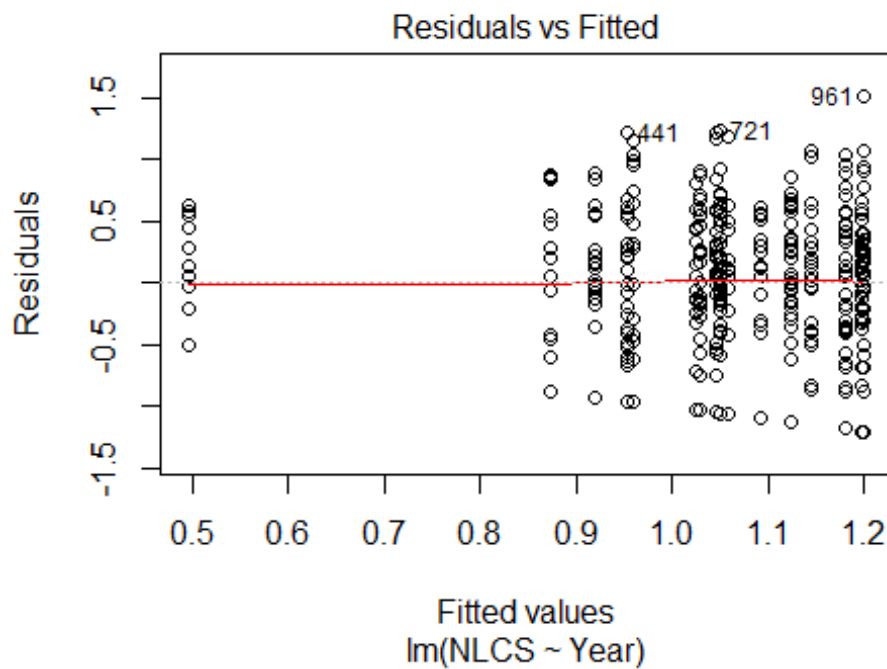
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.424097 -0.324736 0.000915 0.369153 2.029816
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3468 0.5089 2.65 0.0086 **
## LastAuthorFemale1 0.0672 0.0951 0.71 0.4805
## Year1997 -0.1600 0.5237 -0.31 0.7601
## Year1998 -0.4885 0.5311 -0.92 0.3585
## Year1999 -0.1067 0.5358 -0.20 0.8423
## Year2000 -0.3589 0.5218 -0.69 0.4921
## Year2001 -0.4436 0.5427 -0.82 0.4144
## Year2002 -0.6897 0.5627 -1.23 0.2213
## Year2003 0.0773 0.5237 0.15 0.8828
## Year2004 -0.2755 0.5214 -0.53 0.5976
## Year2005 -0.4063 0.5182 -0.78 0.4337
## Year2006 -0.0284 0.5176 -0.05 0.9562
```

```

## Year2007          -0.4409      0.5119   -0.86   0.3898
## Year2008          -0.0778      0.5135   -0.15   0.8797
## Year2009          -0.3770      0.5157   -0.73   0.4653
## Year2010          -0.1266      0.5358   -0.24   0.8134
## Year2011          -0.3389      0.5138   -0.66   0.5101
## Year2012          -0.2898      0.5229   -0.55   0.5799
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.109, Adjusted R-squared:  0.0553
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 277 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0947 0.8750 0.9500 0.9000 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      3.34e-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: 299"
## [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
##   58   69   59   42   73   58   47   33   39   53   45   43   52   70   59
## 2011 2012
##   74   58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   14   19   30   15   25   18   26   14   18   27   17   24   25   33   28
## 2011 2012

```

```
## 35 32
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 15 28 13 24 17 21 14 17 26 17 17 20 26 24
## 2011 2012
## 22 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 = 20, df = 16, p-value = 0.2
```



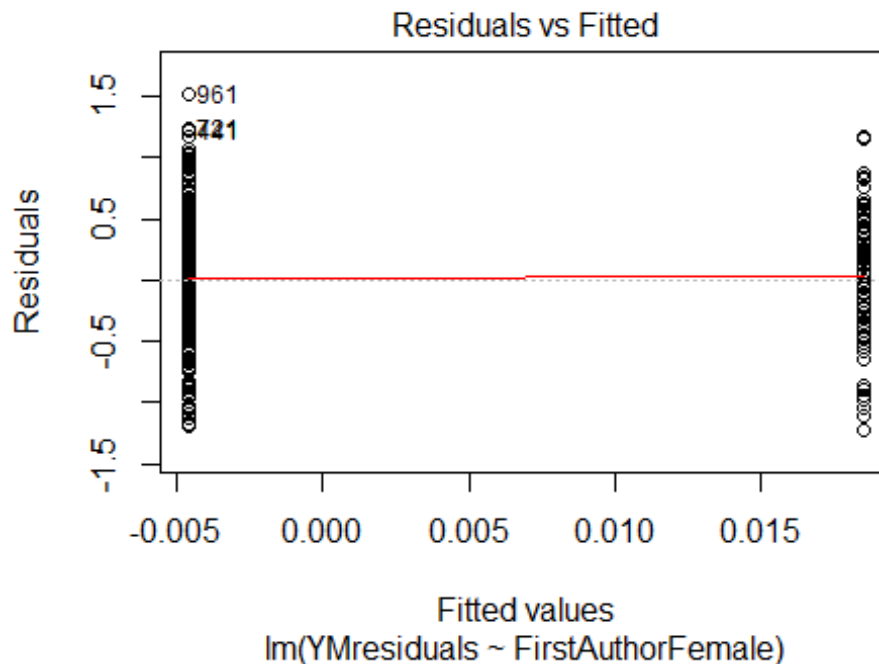
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00011, df = 1, p-value = 1

## [1] "Female first author team size 2018 geometric mean: 2.88506202716261"
## [1] "Male first author team size 2018 geometric mean: 2.54414678130907"

## 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 = 86, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.66709661118172"
## [1] "Male last author team size 2018 geometric mean: 2.64145800488713"

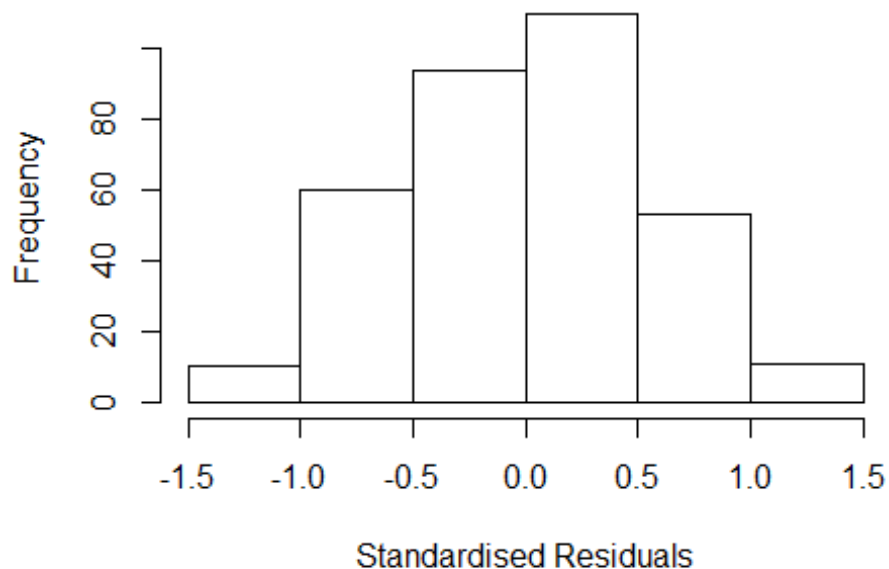
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 71, 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	1.453	1	1.205
LastAuthorFemale	1.349	1	1.161
UniqueAuthors	2.361	4	1.113
Year	2.802	16	1.033

Residuals from first and last author and team size



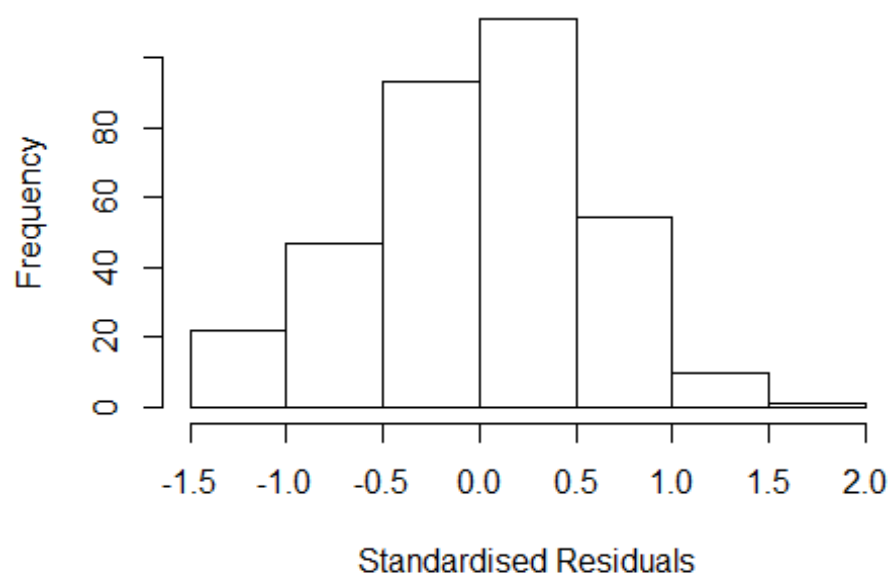
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4132 -0.3630 0.0359 0.3643 1.3954
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89003 0.16771 5.31 2.1e-07 ***
## FirstAuthorFemale1 -0.11349 0.08848 -1.28 0.20056
## LastAuthorFemale1 0.19469 0.09102 2.14 0.03321 *
## UniqueAuthors2 0.29844 0.08447 3.53 0.00047 ***
## UniqueAuthors3 0.24819 0.09442 2.63 0.00900 **
## UniqueAuthors4 0.48733 0.11910 4.09 5.4e-05 ***
## UniqueAuthors5 0.55857 0.14895 3.75 0.00021 ***
## Year1997 -0.20481 0.27269 -0.75 0.45317
## Year1998 -0.01847 0.19375 -0.10 0.92412
## Year1999 -0.22297 0.28591 -0.78 0.43606
```

```

## Year2000      -0.14296      0.19345      -0.74      0.46046
## Year2001      -0.18542      0.20021      -0.93      0.35509
## Year2002      -0.21458      0.19399      -1.11      0.26950
## Year2003      -0.13537      0.21520      -0.63      0.52979
## Year2004       0.02761      0.21242       0.13      0.89665
## Year2005      -0.12282      0.18709      -0.66      0.51198
## Year2006      -0.56559      0.18882      -3.00      0.00296 **
## Year2007      -0.06668      0.23361      -0.29      0.77549
## Year2008      -0.01494      0.20072      -0.07      0.94070
## Year2009       0.03582      0.19591       0.18      0.85503
## Year2010       0.05963      0.22249       0.27      0.78886
## Year2011      -0.15639      0.20560      -0.76      0.44742
## Year2012       0.00972      0.18328       0.05      0.95775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.575
## Multiple R-squared:  0.16,   Adjusted R-squared:  0.101
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 310 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.526  0.865   0.955   0.915   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           2.96e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "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  1.347 1           1.161
## Year              1.276 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
## -1.3392 -0.3621 0.0267 0.3954 1.5218
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09309 0.15976 6.84 4e-11 ***
## FirstAuthorFemale1 -0.03882 0.08874 -0.44 0.6621
## LastAuthorFemale1 0.18882 0.09352 2.02 0.0443 *
## Year1997 -0.28142 0.29948 -0.94 0.3481
## Year1998 -0.02937 0.20971 -0.14 0.8887
## Year1999 -0.22130 0.27004 -0.82 0.4131
## Year2000 -0.17123 0.19009 -0.90 0.3684
## Year2001 -0.13227 0.19867 -0.67 0.5060
## Year2002 -0.25409 0.19919 -1.28 0.2030
## Year2003 -0.03304 0.22997 -0.14 0.8859
## Year2004 0.01981 0.20327 0.10 0.9224
## Year2005 -0.09217 0.20515 -0.45 0.6535
```

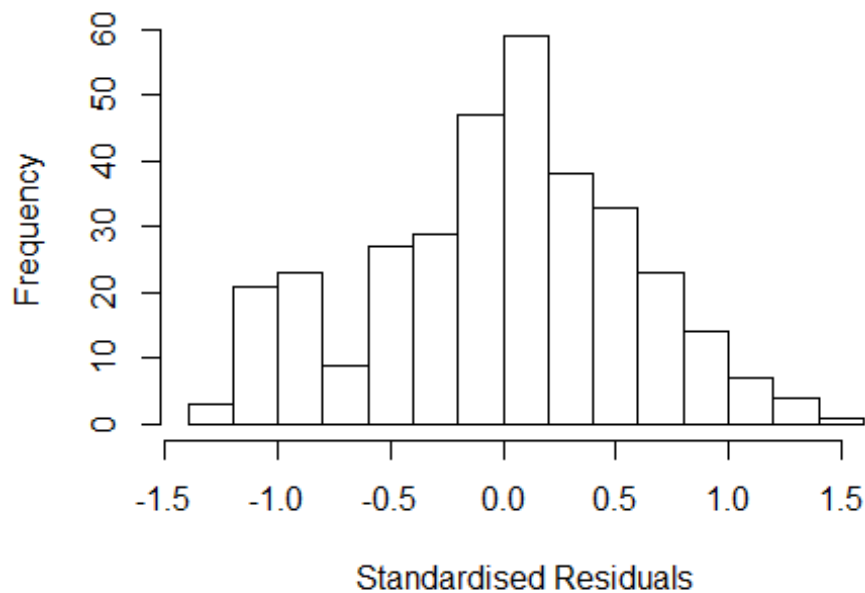


```

## Year2006          -0.60567      0.18760      -3.23      0.0014 **
## Year2007          -0.07165      0.24043      -0.30      0.7659
## Year2008          -0.00317      0.20399      -0.02      0.9876
## Year2009           0.00738      0.19491       0.04      0.9698
## Year2010           0.09609      0.21572       0.45      0.6563
## Year2011          -0.16573      0.21295      -0.78      0.4370
## Year2012           0.07226      0.18234       0.40      0.6921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.078, Adjusted R-squared:  0.026
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 306 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.473  0.856  0.951  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      2.96e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.115 1      1.056
## Year              1.115 16      1.003

```

Residuals from first author



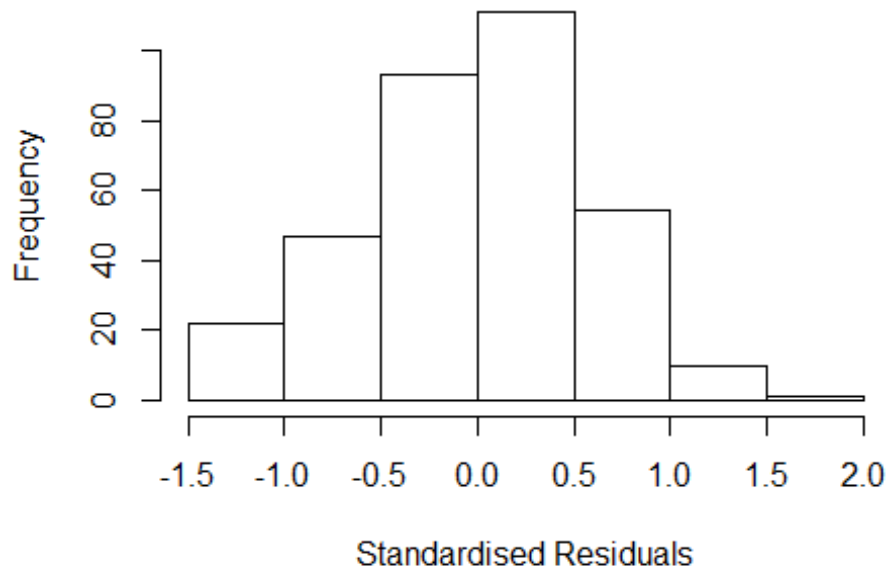
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.223 -0.389 0.039 0.388 1.504
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0841 0.1590 6.82 4.6e-11 ***
## FirstAuthorFemale1 0.0160 0.0804 0.20 0.8424
## Year1997 -0.2268 0.3168 -0.72 0.4744
## Year1998 -0.0191 0.2086 -0.09 0.9272
## Year1999 -0.1986 0.2730 -0.73 0.4675
## Year2000 -0.1476 0.1902 -0.78 0.4385
## Year2001 -0.0859 0.1974 -0.44 0.6637
## Year2002 -0.2037 0.1980 -1.03 0.3042
## Year2003 -0.0201 0.2257 -0.09 0.9291
## Year2004 0.0263 0.2018 0.13 0.8966
## Year2005 -0.0760 0.2030 -0.37 0.7085
## Year2006 -0.5949 0.1863 -3.19 0.0015 **
```

```

## Year2007          -0.0253      0.2294   -0.11   0.9124
## Year2008           0.0365      0.2015    0.18   0.8564
## Year2009           0.0294      0.1955    0.15   0.8804
## Year2010           0.1229      0.2190    0.56   0.5752
## Year2011          -0.1293      0.2106   -0.61   0.5396
## Year2012           0.0975      0.1829    0.53   0.5944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.581
## Multiple R-squared:  0.0673, Adjusted R-squared:  0.0178
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 307 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.482  0.848  0.949  0.906  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      2.96e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



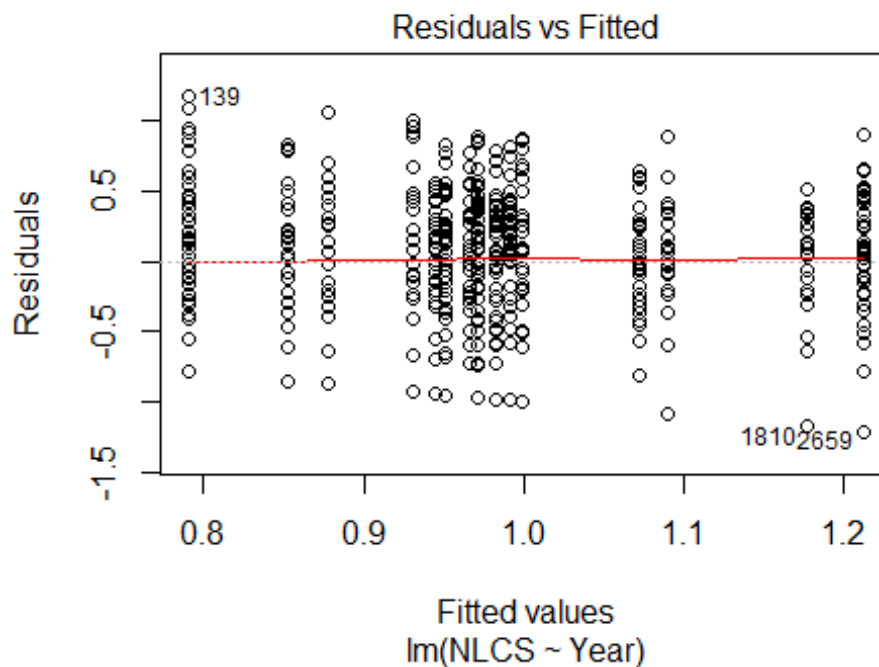
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3621 -0.3594 0.0308 0.3925 1.5242
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08701 0.15810 6.88 3.2e-11 ***
## LastAuthorFemale1 0.17525 0.08651 2.03 0.0436 *
## Year1997 -0.27722 0.30276 -0.92 0.3605
## Year1998 -0.02819 0.20868 -0.14 0.8926
## Year1999 -0.22656 0.26993 -0.84 0.4019
## Year2000 -0.17049 0.19005 -0.90 0.3703
## Year2001 -0.12748 0.19739 -0.65 0.5188
## Year2002 -0.25076 0.19846 -1.26 0.2073
## Year2003 -0.03551 0.22797 -0.16 0.8763
## Year2004 0.00937 0.20197 0.05 0.9630
## Year2005 -0.09386 0.20580 -0.46 0.6486
## Year2006 -0.60583 0.18775 -3.23 0.0014 **
```

```

## Year2007          -0.06870      0.23971    -0.29    0.7746
## Year2008          -0.00350      0.20360    -0.02    0.9863
## Year2009           0.00950      0.19458     0.05    0.9611
## Year2010           0.09981      0.21527     0.46    0.6432
## Year2011          -0.16162      0.21238    -0.76    0.4472
## Year2012           0.07271      0.18231     0.40    0.6903
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0775, Adjusted R-squared:  0.0285
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 308 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.468  0.856  0.950  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      2.96e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 338"
## [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
##   94   75   93   67   66   77   68   56   87   59   86   75   76   78   99
## 2011 2012
##  108  126
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   28   45   27   30   30   26   23   36   26   32   32   23   32   41
## 2011 2012

```

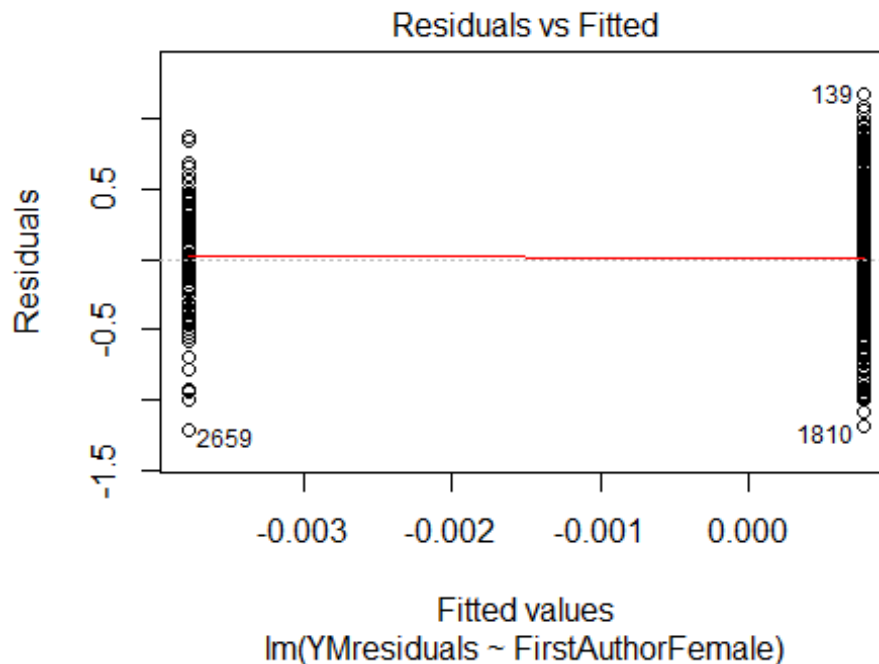
```
## 45 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 24 36 20 29 22 26 20 33 21 27 25 19 27 38
## 2011 2012
## 37 46
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.09
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.9, df = 1, p-value = 0.05
## [1] "Female first author team size 2018 geometric mean: 2.88954547483531"
## [1] "Male first author team size 2018 geometric mean: 2.36175342481457"
## 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 = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.60517108469735"
## [1] "Male last author team size 2018 geometric mean: 2.42734176515744"

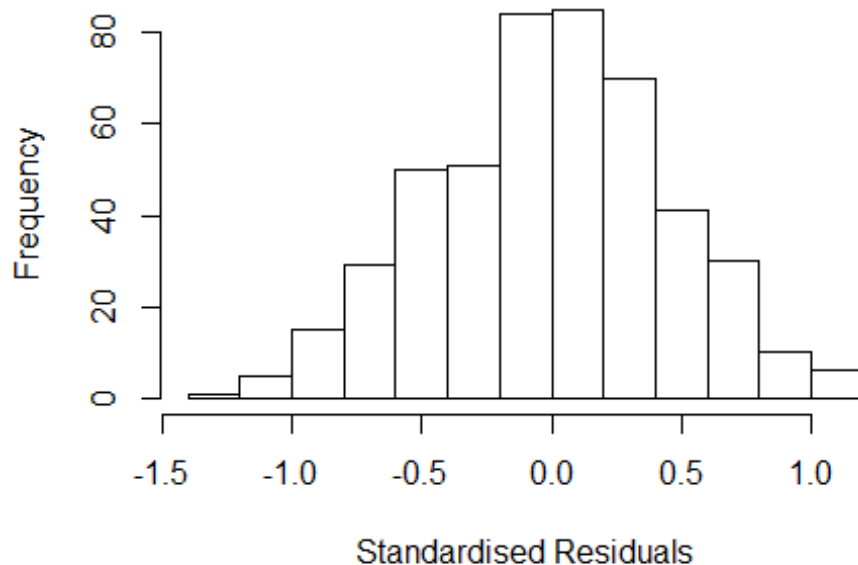
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.8
## 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	1.273	1	1.128
UniqueAuthors	2.267	4	1.108
Year	2.567	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
## -1.20833 -0.31676 0.00324 0.31446 1.18224
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6247 0.1247 5.01 7.8e-07 ***
## FirstAuthorFemale1 -0.0816 0.0535 -1.52 0.128
## LastAuthorFemale1 0.0269 0.0653 0.41 0.681
## UniqueAuthors2 0.3087 0.0596 5.18 3.3e-07 ***
## UniqueAuthors3 0.3358 0.0675 4.98 9.2e-07 ***
## UniqueAuthors4 0.3142 0.0776 4.05 6.1e-05 ***
## UniqueAuthors5 0.1962 0.0780 2.51 0.012 *
## Year1997 0.1573 0.1774 0.89 0.375
## Year1998 0.2387 0.1447 1.65 0.100 .
## Year1999 0.1778 0.1499 1.19 0.236
```

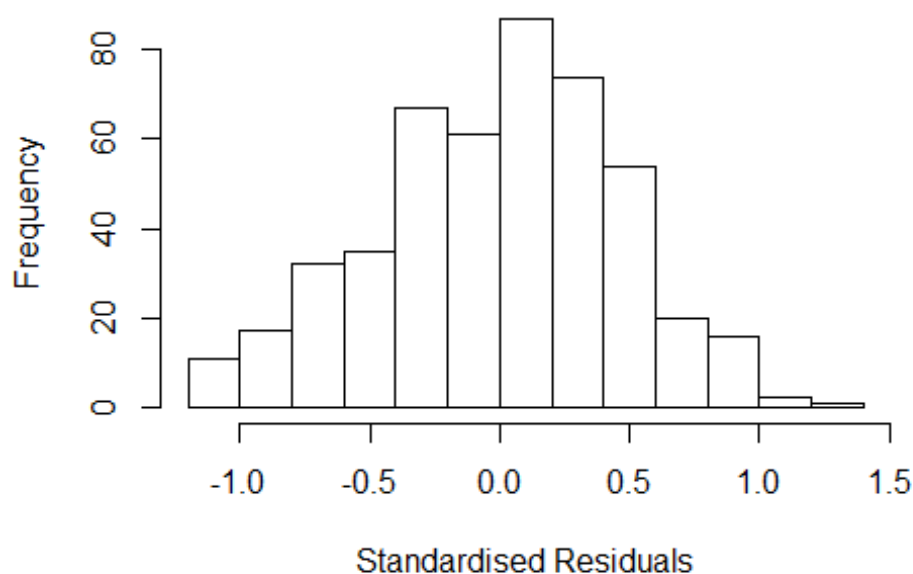


```

## Year2000          -0.0669      0.1540    -0.43      0.664
## Year2001           0.0375      0.1632      0.23      0.819
## Year2002           0.0852      0.1656      0.51      0.607
## Year2003           0.0859      0.1632      0.53      0.599
## Year2004           0.1878      0.1519      1.24      0.217
## Year2005           0.2749      0.1499      1.83      0.067 .
## Year2006           0.0864      0.1466      0.59      0.556
## Year2007           0.2986      0.1833      1.63      0.104
## Year2008           0.3466      0.1530      2.27      0.024 *
## Year2009           0.1844      0.1399      1.32      0.188
## Year2010           0.1880      0.1485      1.27      0.206
## Year2011           0.1308      0.1423      0.92      0.358
## Year2012           0.3562      0.1445      2.46      0.014 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.47
## Multiple R-squared:  0.151, Adjusted R-squared:  0.11
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 425 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.488  0.869  0.948  0.912  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          2.10e-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 1.221 1 1.105
## LastAuthorFemale 1.190 1 1.091
## Year 1.308 16 1.008

```

Residuals from first and last author



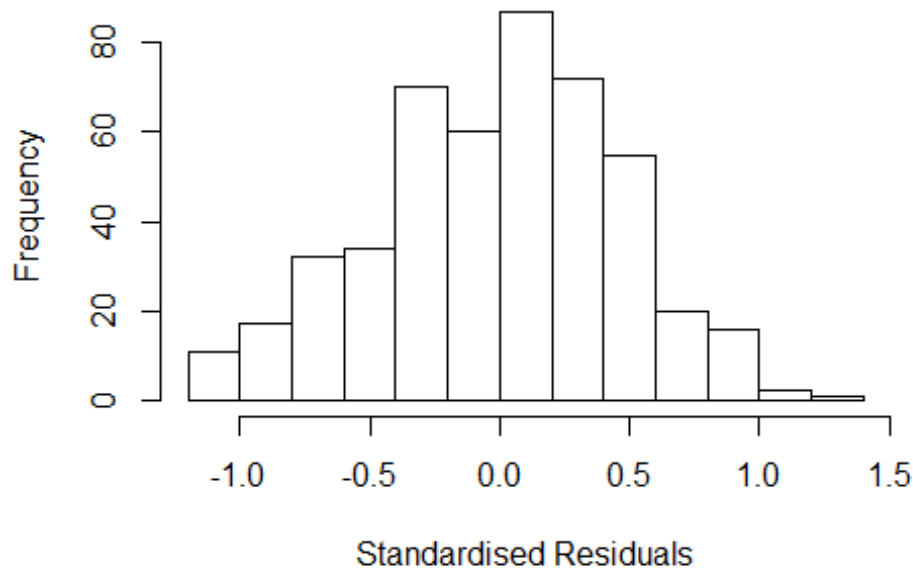
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1912 -0.3019  0.0365  0.3237  1.2401
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6930    0.1285    5.39 1.1e-07 ***
## FirstAuthorFemale1 -0.0323    0.0527   -0.61 0.54012
## LastAuthorFemale1  0.0349    0.0621    0.56 0.57479
## Year1997         0.2002    0.1858    1.08 0.28174
## Year1998         0.2522    0.1487    1.70 0.09059 .
## Year1999         0.2693    0.1588    1.70 0.09059 .
## Year2000         0.0674    0.1630    0.41 0.67930
## Year2001         0.1108    0.1735    0.64 0.52327
## Year2002         0.1947    0.1678    1.16 0.24649
## Year2003         0.2939    0.1670    1.76 0.07910 .
## Year2004         0.2949    0.1699    1.74 0.08327 .
## Year2005         0.4095    0.1504    2.72 0.00670 **
```

```

## Year2006          0.2657      0.1505      1.77  0.07807 .
## Year2007          0.3599      0.1979      1.82  0.06958 .
## Year2008          0.4982      0.1520      3.28  0.00113 **
## Year2009          0.3240      0.1413      2.29  0.02230 *
## Year2010          0.3299      0.1498      2.20  0.02816 *
## Year2011          0.2295      0.1479      1.55  0.12131
## Year2012          0.5201      0.1447      3.59  0.00036 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.0731, Adjusted R-squared:  0.0367
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 433 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.491  0.873   0.951   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      2.10e-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 1.157 1      1.076
## Year              1.157 16      1.005

```

Residuals from first author



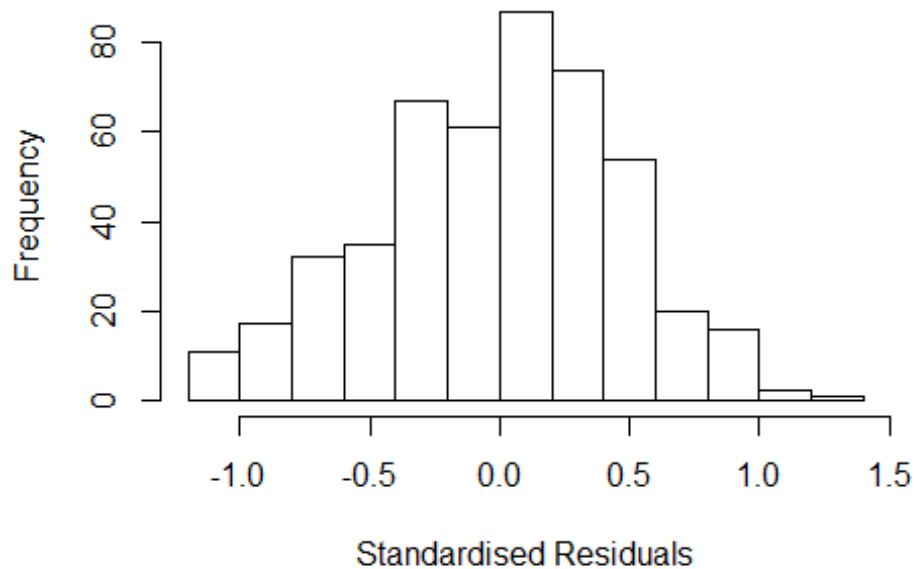
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1976 -0.3084 0.0364 0.3361 1.2725
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6955 0.1283 5.42 9.6e-08 ***
## FirstAuthorFemale1 -0.0221 0.0514 -0.43 0.66731
## Year1997 0.2002 0.1851 1.08 0.28016
## Year1998 0.2524 0.1484 1.70 0.08957 .
## Year1999 0.2654 0.1580 1.68 0.09371 .
## Year2000 0.0726 0.1618 0.45 0.65408
## Year2001 0.1140 0.1728 0.66 0.50967
## Year2002 0.1950 0.1673 1.17 0.24442
## Year2003 0.2918 0.1671 1.75 0.08137 .
## Year2004 0.2932 0.1699 1.73 0.08505 .
## Year2005 0.4111 0.1495 2.75 0.00619 **
## Year2006 0.2645 0.1503 1.76 0.07909 .
```

```

## Year2007          0.3606      0.1982      1.82  0.06955 .
## Year2008          0.5021      0.1507      3.33  0.00094 ***
## Year2009          0.3218      0.1412      2.28  0.02312 *
## Year2010          0.3294      0.1494      2.20  0.02797 *
## Year2011          0.2330      0.1469      1.59  0.11351
## Year2012          0.5183      0.1446      3.59  0.00037 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.0728, Adjusted R-squared:  0.0384
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.468  0.876  0.952  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      2.10e-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 1.134 1      1.065
## Year              1.134 16      1.004

```

Residuals from last author



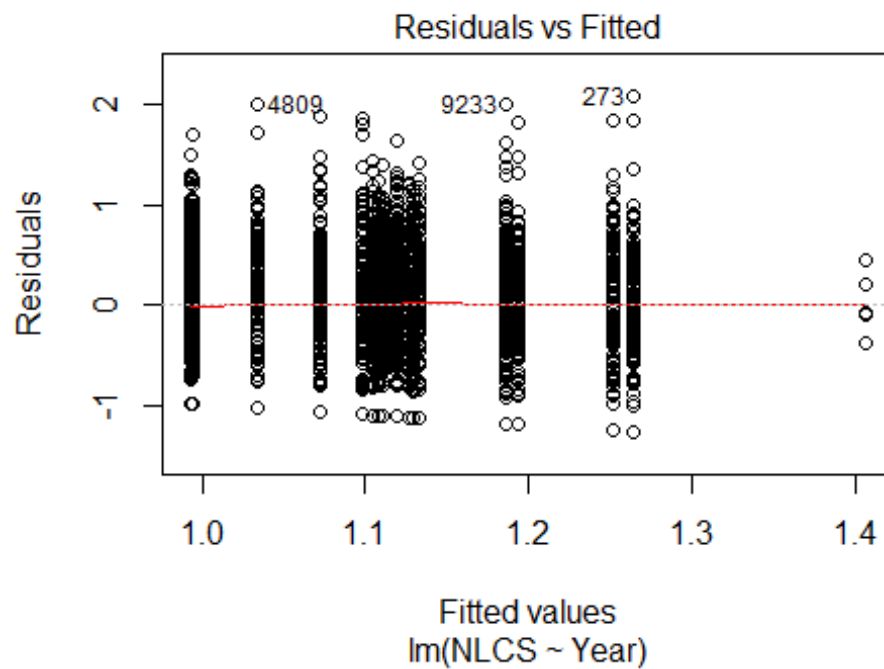
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.205 -0.306  0.032  0.329  1.255
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6900    0.1288   5.36 1.3e-07 ***
## LastAuthorFemale1 0.0228    0.0612   0.37 0.70949
## Year1997        0.2038    0.1859   1.10 0.27373
## Year1998        0.2527    0.1493   1.69 0.09123 .
## Year1999        0.2650    0.1595   1.66 0.09730 .
## Year2000        0.0673    0.1638   0.41 0.68135
## Year2001        0.1141    0.1737   0.66 0.51150
## Year2002        0.1891    0.1676   1.13 0.25977
## Year2003        0.2921    0.1678   1.74 0.08249 .
## Year2004        0.2923    0.1701   1.72 0.08643 .
## Year2005        0.4093    0.1508   2.71 0.00689 **
## Year2006        0.2598    0.1509   1.72 0.08585 .
```

```

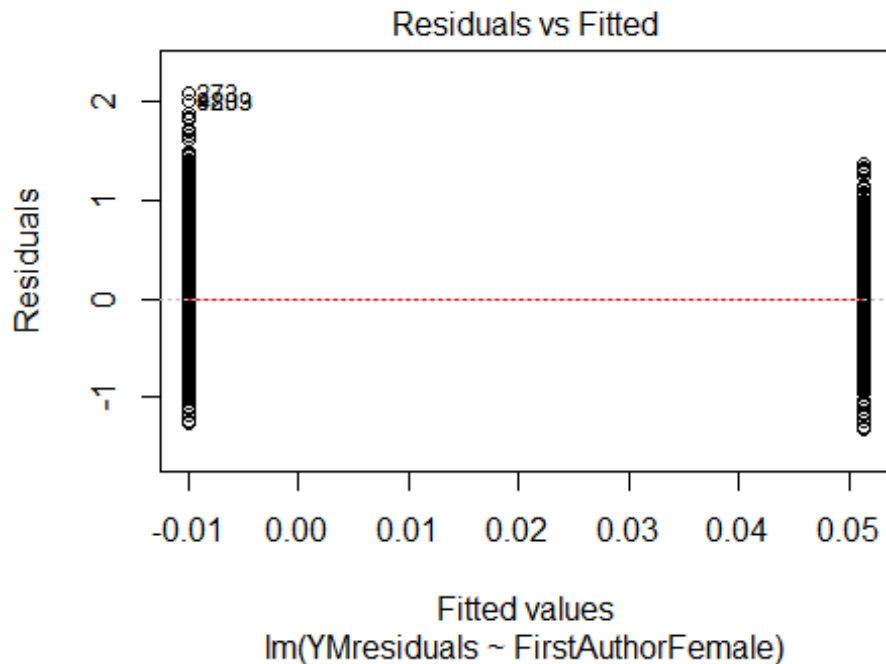
## Year2007          0.3610      0.1985      1.82  0.06956 .
## Year2008          0.5000      0.1525      3.28  0.00112 **
## Year2009          0.3204      0.1420      2.26  0.02452 *
## Year2010          0.3310      0.1503      2.20  0.02812 *
## Year2011          0.2311      0.1486      1.56  0.12050
## Year2012          0.5155      0.1457      3.54  0.00044 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.0729, Adjusted R-squared:  0.0385
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 439 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.479  0.875   0.953   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      2.10e-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: 477"
## [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
##  542  524  560   37  545  586  459  495  457  429  550  528  578  663  671
## 2011 2012
##  726  690
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  131  114  113    6  150  129  173  161  175  172  240  219  234  305  314
## 2011 2012

```

```
## 325 329
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 105 93 87 6 121 112 154 139 149 136 206 188 201 243 258
## 2011 2012
## 253 267
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 49, df = 16, p-value = 4e-05
```

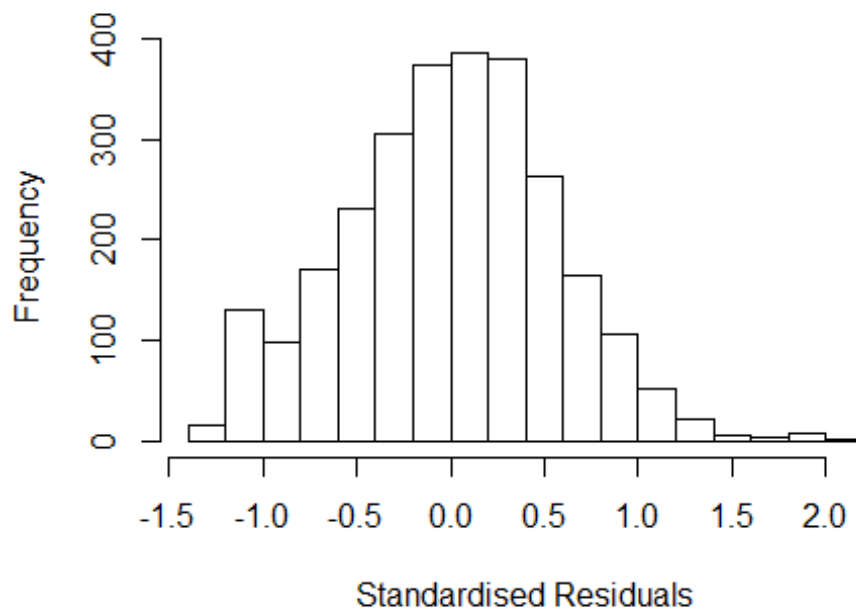


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.8, df = 1, p-value = 0.002
```

```
## [1] "Female first author team size 2018 geometric mean: 4.03894655743926"
## [1] "Male first author team size 2018 geometric mean: 3.18996232416878"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.04999918761414"
## [1] "Male last author team size 2018 geometric mean: 3.21974230511602"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, 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 1.071 1          1.035
## LastAuthorFemale  1.082 1          1.040
## UniqueAuthors    1.239 4          1.027
## Year              1.229 16         1.006
```

Residuals from first and last author and team size



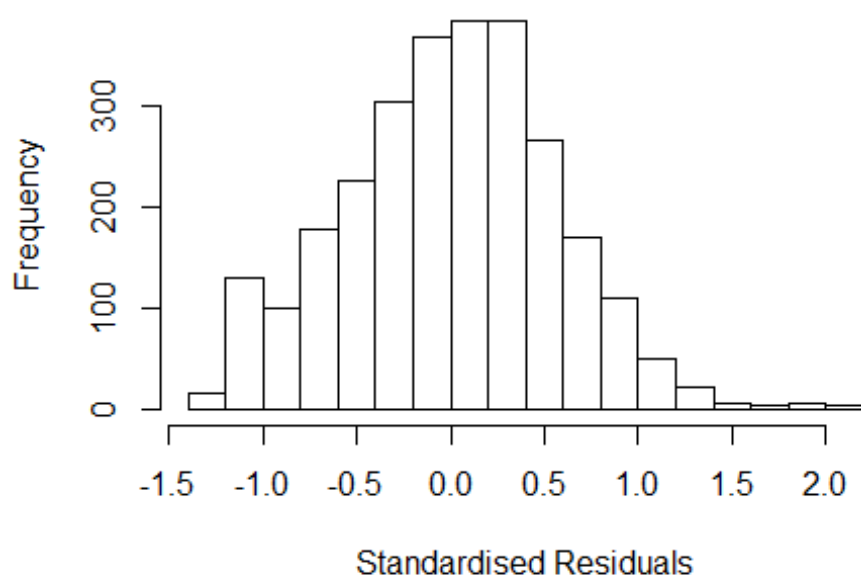
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3483 -0.3736 0.0205 0.3712 2.1157
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2283 0.0657 18.70 < 2e-16 ***
## FirstAuthorFemale1 0.0478 0.0277 1.72 0.08475 .
## LastAuthorFemale1 -0.0228 0.0370 -0.62 0.53798
## UniqueAuthors2 0.0933 0.0406 2.30 0.02160 *
## UniqueAuthors3 0.0409 0.0402 1.02 0.30905
## UniqueAuthors4 0.0567 0.0429 1.32 0.18702
## UniqueAuthors5 0.1428 0.0417 3.42 0.00064 ***
## Year1997 -0.0427 0.0897 -0.48 0.63401
## Year1998 -0.1523 0.0919 -1.66 0.09749 .
## Year1999 0.0721 0.1227 0.59 0.55681
```

```

## Year2000          -0.1201      0.0786   -1.53   0.12677
## Year2001          -0.1504      0.0913   -1.65   0.09946 .
## Year2002          -0.2018      0.0843   -2.39   0.01676 *
## Year2003          -0.2988      0.0819   -3.65   0.00027 ***
## Year2004          -0.3216      0.0811   -3.96   7.6e-05 ***
## Year2005          -0.1931      0.0764   -2.53   0.01149 *
## Year2006          -0.3258      0.0716   -4.55   5.7e-06 ***
## Year2007          -0.2049      0.0770   -2.66   0.00785 **
## Year2008          -0.2473      0.0727   -3.40   0.00068 ***
## Year2009          -0.1221      0.0702   -1.74   0.08202 .
## Year2010          -0.1957      0.0689   -2.84   0.00456 **
## Year2011          -0.1802      0.0689   -2.62   0.00896 **
## Year2012          -0.1884      0.0682   -2.76   0.00580 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.027, Adjusted R-squared:  0.0191
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2498 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.860  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          3.68e-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 1.040 1          1.020
## LastAuthorFemale  1.049 1          1.024
## Year              1.067 16          1.002

```

Residuals from first and last author



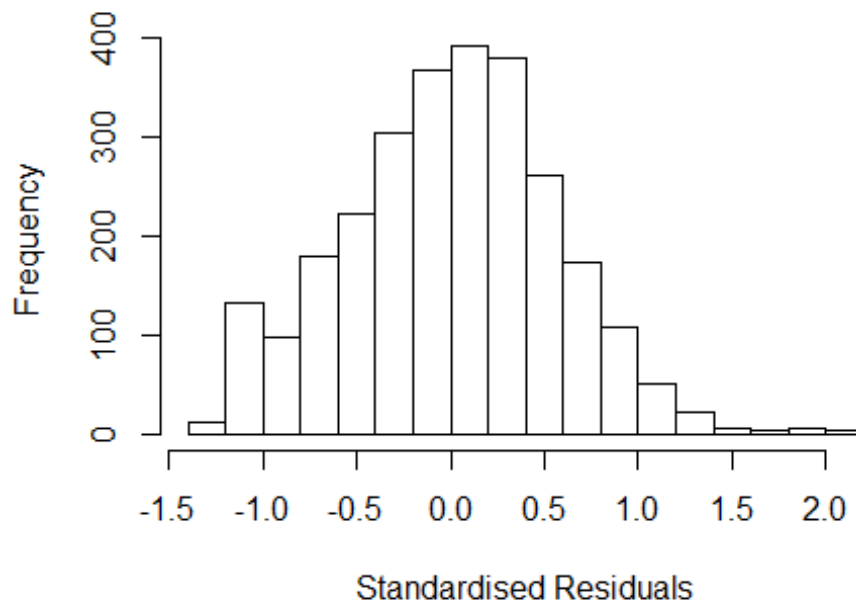
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3268 -0.3717 0.0226 0.3726 2.0611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2829 0.0584 21.96 < 2e-16 ***
## FirstAuthorFemale1 0.0635 0.0275 2.31 0.02118 *
## LastAuthorFemale1 -0.0196 0.0367 -0.53 0.59362
## Year1997 -0.0377 0.0883 -0.43 0.66909
## Year1998 -0.1461 0.0915 -1.60 0.11033
## Year1999 0.1243 0.1249 1.00 0.31951
## Year2000 -0.1147 0.0772 -1.49 0.13754
## Year2001 -0.1484 0.0901 -1.65 0.09969 .
## Year2002 -0.1903 0.0832 -2.29 0.02228 *
## Year2003 -0.2932 0.0807 -3.63 0.00028 ***
## Year2004 -0.3221 0.0800 -4.03 5.8e-05 ***
## Year2005 -0.1800 0.0749 -2.40 0.01634 *
```

```

## Year2006          -0.3210      0.0700   -4.58  4.8e-06 ***
## Year2007          -0.1932      0.0755   -2.56  0.01056 *
## Year2008          -0.2345      0.0711   -3.30  0.00100 ***
## Year2009          -0.1133      0.0685   -1.65  0.09843 .
## Year2010          -0.1801      0.0670   -2.69  0.00721 **
## Year2011          -0.1653      0.0673   -2.46  0.01410 *
## Year2012          -0.1693      0.0662   -2.55  0.01067 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0211, Adjusted R-squared:  0.0145
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 207 weights are ~= 1. The remaining 2511 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.138  0.867  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      3.68e-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 1.029 1      1.014
## Year      1.029 16      1.001

```

Residuals from first author



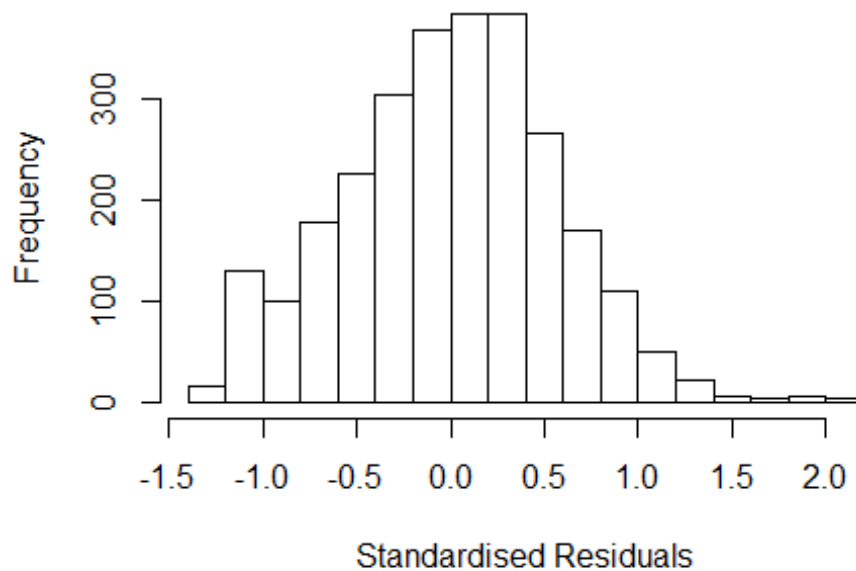
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3436 -0.3717 0.0232 0.3744 2.0620
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2820 0.0585 21.93 < 2e-16 ***
## FirstAuthorFemale1 0.0616 0.0274 2.25 0.02445 *
## Year1997 -0.0375 0.0883 -0.42 0.67108
## Year1998 -0.1481 0.0913 -1.62 0.10483
## Year1999 0.1218 0.1236 0.99 0.32438
## Year2000 -0.1146 0.0773 -1.48 0.13826
## Year2001 -0.1490 0.0901 -1.65 0.09812 .
## Year2002 -0.1916 0.0831 -2.31 0.02121 *
## Year2003 -0.2933 0.0807 -3.63 0.00028 ***
## Year2004 -0.3228 0.0799 -4.04 5.5e-05 ***
## Year2005 -0.1818 0.0747 -2.43 0.01508 *
## Year2006 -0.3221 0.0699 -4.61 4.3e-06 ***
```

```

## Year2007          -0.1944      0.0754   -2.58  0.00999 **
## Year2008          -0.2351      0.0712   -3.30  0.00097 ***
## Year2009          -0.1143      0.0685   -1.67  0.09529 .
## Year2010          -0.1813      0.0669   -2.71  0.00676 **
## Year2011          -0.1664      0.0672   -2.48  0.01331 *
## Year2012          -0.1709      0.0661   -2.59  0.00975 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.021, Adjusted R-squared:  0.0148
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 214 weights are ~= 1. The remaining 2504 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.866  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      3.68e-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 1.037 1          1.019
## 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
## -1.2886 -0.3751 0.0188 0.3723 2.0554
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2886 0.0584 22.07 < 2e-16 ***
## LastAuthorFemale1 -0.0104 0.0368 -0.28 0.77775
## Year1997 -0.0378 0.0887 -0.43 0.66963
## Year1998 -0.1426 0.0919 -1.55 0.12084
## Year1999 0.1169 0.1243 0.94 0.34690
## Year2000 -0.1092 0.0769 -1.42 0.15576
## Year2001 -0.1477 0.0904 -1.63 0.10226
## Year2002 -0.1861 0.0829 -2.24 0.02487 *
## Year2003 -0.2888 0.0806 -3.58 0.00034 ***
## Year2004 -0.3206 0.0800 -4.01 6.3e-05 ***
## Year2005 -0.1767 0.0749 -2.36 0.01833 *
## Year2006 -0.3185 0.0702 -4.54 6.0e-06 ***
```

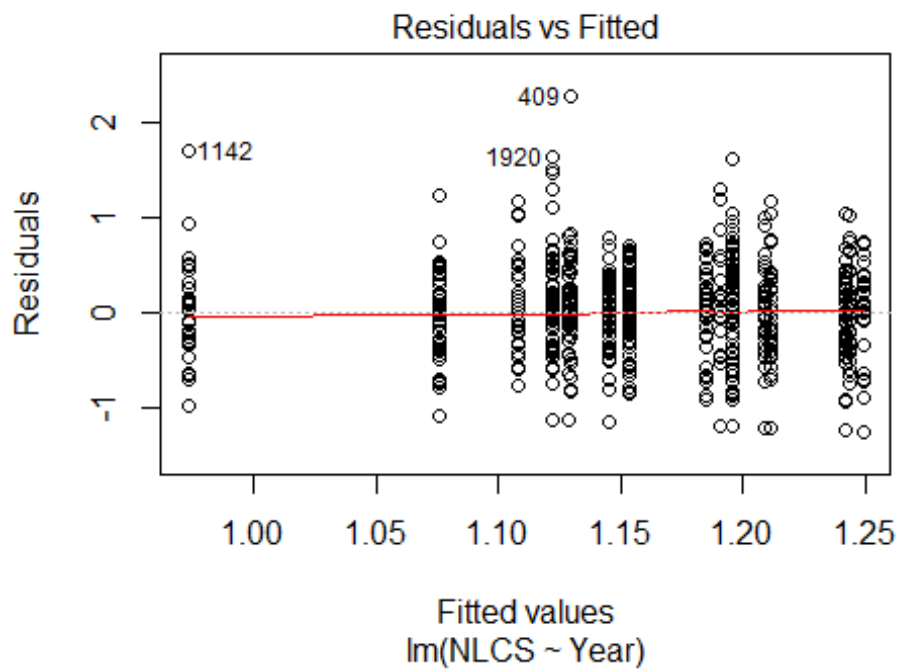


```

## Year2007          -0.1881      0.0756   -2.49  0.01284 *
## Year2008          -0.2278      0.0712   -3.20  0.00140 **
## Year2009          -0.1072      0.0684   -1.57  0.11737
## Year2010          -0.1738      0.0670   -2.60  0.00951 **
## Year2011          -0.1586      0.0672   -2.36  0.01830 *
## Year2012          -0.1634      0.0661   -2.47  0.01348 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0194, Adjusted R-squared:  0.0132
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 221 weights are ~= 1. The remaining 2497 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.141  0.866  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      3.68e-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: 2718"
## [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
##   93  109   92  105  123   84   62   62   74   68   78   89   96  126   84
## 2011 2012
##  109  116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   33   31   36   47   23   27   33   46   34   36   36   37   58   46
## 2011 2012

```

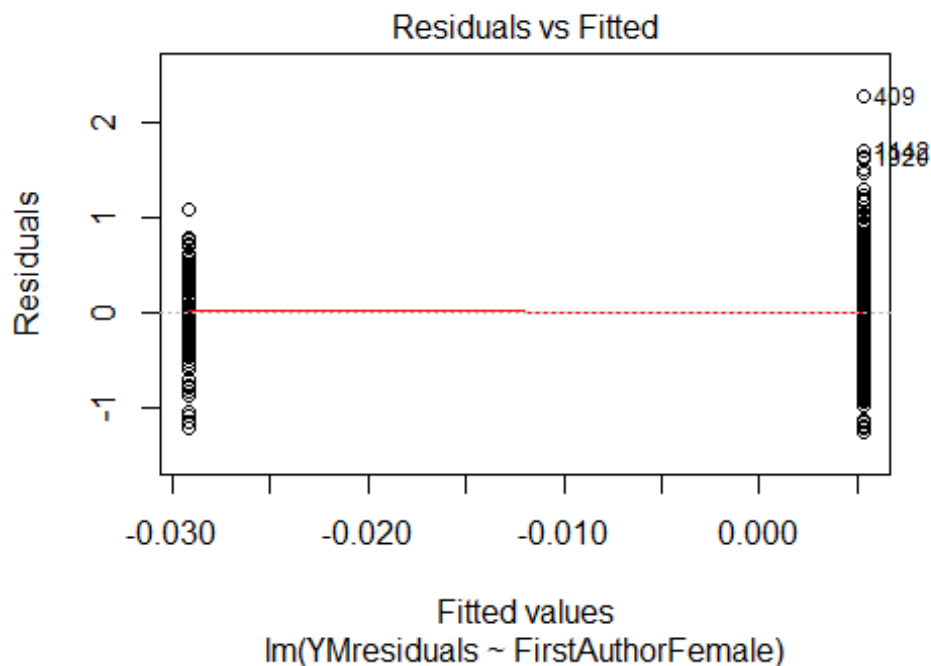
```
## 67 67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 28 28 29 39 20 24 26 41 30 30 32 30 47 38
## 2011 2012
## 57 45
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.008
```



```
##
## 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: 3.99623710216289"
## [1] "Male first author team size 2018 geometric mean: 4.01854803384126"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

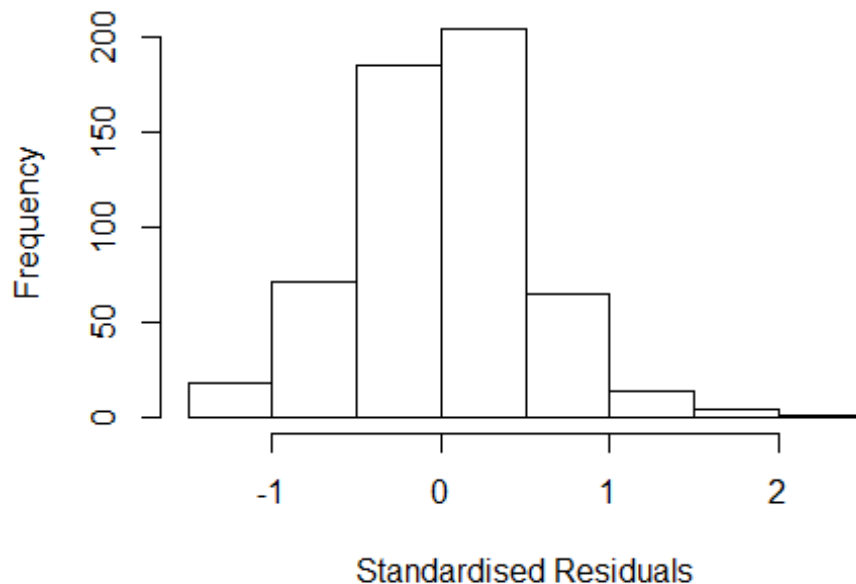
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 290, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.38433904700863"
## [1] "Male last author team size 2018 geometric mean: 3.91938308865561"

## 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 = 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 1.242  1      1.115
## LastAuthorFemale  1.216  1      1.102
## UniqueAuthors    1.989  4      1.090
## Year              2.486 16      1.029
```

Residuals from first and last author and team size



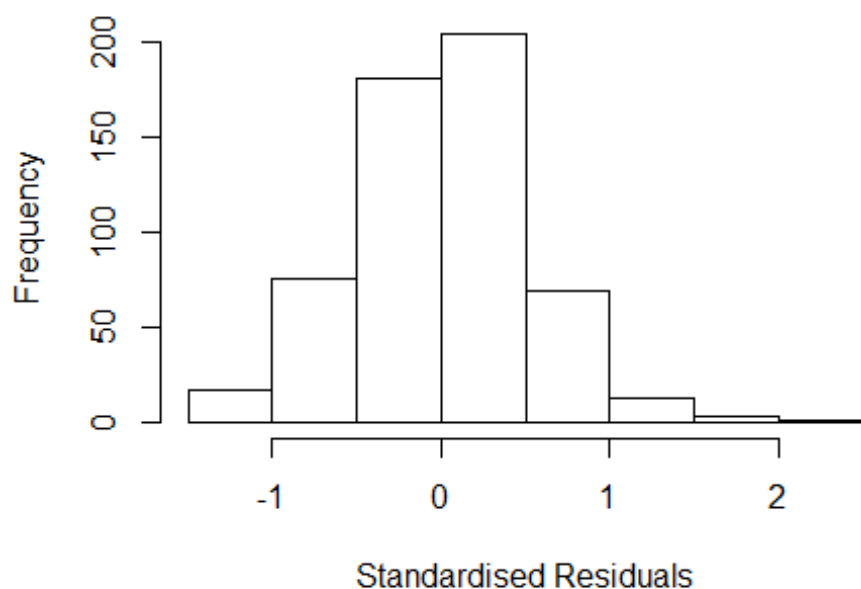
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3120 -0.3089 0.0149 0.3321 2.2501
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16247 0.23672 4.91 1.2e-06 ***
## FirstAuthorFemale1 -0.04120 0.05589 -0.74 0.46
## LastAuthorFemale1 0.00362 0.05671 0.06 0.95
## UniqueAuthors2 0.04666 0.07630 0.61 0.54
## UniqueAuthors3 0.09317 0.07603 1.23 0.22
## UniqueAuthors4 0.09054 0.08314 1.09 0.28
## UniqueAuthors5 0.09760 0.08081 1.21 0.23
## Year1997 0.06228 0.24365 0.26 0.80
## Year1998 0.07063 0.23891 0.30 0.77
## Year1999 -0.10072 0.24814 -0.41 0.68
```

```

## Year2000      -0.03323    0.23299   -0.14    0.89
## Year2001      -0.02811    0.28490   -0.10    0.92
## Year2002       0.05895    0.24002    0.25    0.81
## Year2003      -0.02219    0.23709   -0.09    0.93
## Year2004       0.05241    0.22509    0.23    0.82
## Year2005      -0.09314    0.22683   -0.41    0.68
## Year2006      -0.31592    0.23315   -1.35    0.18
## Year2007      -0.13087    0.23550   -0.56    0.58
## Year2008      -0.06922    0.23650   -0.29    0.77
## Year2009      -0.03732    0.23122   -0.16    0.87
## Year2010      -0.13605    0.22713   -0.60    0.55
## Year2011      -0.05057    0.22487   -0.22    0.82
## Year2012      -0.06528    0.23742   -0.27    0.78
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.47
## Multiple R-squared:  0.0338, Adjusted R-squared:  -0.00566
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 88 is an outlier with |weight| = 0 ( < 0.00018);
## 42 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.118  0.862  0.947   0.896  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           1.78e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
## factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.182 1 1.087
## LastAuthorFemale 1.149 1 1.072
## Year 1.359 16 1.010

```

Residuals from first and last author



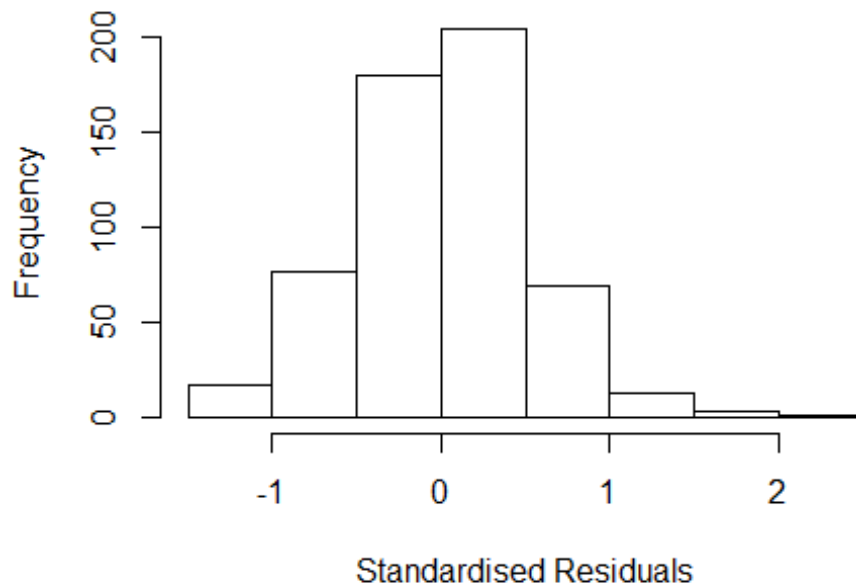
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2741 -0.3224 0.0202 0.3261 2.2925
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2250 0.2169 5.65 2.6e-08 ***
## FirstAuthorFemale1 -0.0303 0.0549 -0.55 0.58
## LastAuthorFemale1 0.0041 0.0556 0.07 0.94
## Year1997 0.0492 0.2394 0.21 0.84
## Year1998 0.0591 0.2355 0.25 0.80
## Year1999 -0.1125 0.2484 -0.45 0.65
## Year2000 -0.0476 0.2294 -0.21 0.84
## Year2001 -0.0375 0.2800 -0.13 0.89
## Year2002 0.0408 0.2407 0.17 0.87
## Year2003 -0.0199 0.2375 -0.08 0.93
## Year2004 0.0567 0.2246 0.25 0.80
## Year2005 -0.0735 0.2259 -0.33 0.74
```

```

## Year2006          -0.3146      0.2332   -1.35      0.18
## Year2007          -0.1261      0.2365   -0.53      0.59
## Year2008          -0.0640      0.2360   -0.27      0.79
## Year2009          -0.0366      0.2304   -0.16      0.87
## Year2010          -0.1282      0.2286   -0.56      0.58
## Year2011          -0.0457      0.2254   -0.20      0.84
## Year2012          -0.0435      0.2357   -0.18      0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.47
## Multiple R-squared:  0.0292, Adjusted R-squared:  -0.00303
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 88 is an outlier with |weight| = 0 ( < 0.00018);
## 42 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.126  0.858  0.950   0.896  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          1.78e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.181  1           1.087
## Year              1.181 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
## -1.2745 -0.3228  0.0196  0.3258  2.2922
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2254     0.2170   5.65 2.7e-08 ***
## FirstAuthorFemale1 -0.0299     0.0551  -0.54   0.59
## Year1997          0.0491     0.2395   0.21   0.84
## Year1998          0.0591     0.2354   0.25   0.80
## Year1999         -0.1126     0.2486  -0.45   0.65
## Year2000         -0.0471     0.2290  -0.21   0.84
## Year2001         -0.0372     0.2800  -0.13   0.89
## Year2002          0.0412     0.2405   0.17   0.86
## Year2003         -0.0203     0.2377  -0.09   0.93
## Year2004          0.0566     0.2246   0.25   0.80
## Year2005         -0.0734     0.2258  -0.32   0.75
## Year2006         -0.3151     0.2333  -1.35   0.18
```

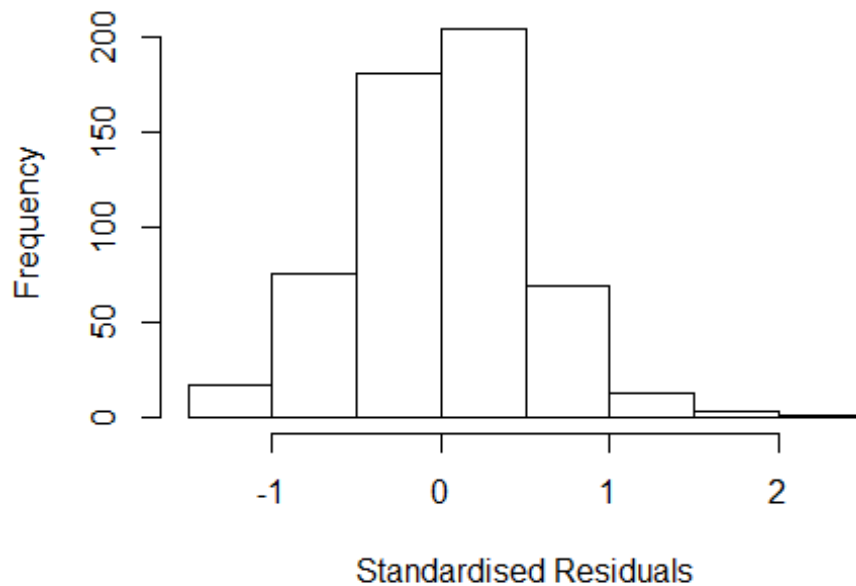


```

## Year2007          -0.1262      0.2364   -0.53      0.59
## Year2008          -0.0640      0.2359   -0.27      0.79
## Year2009          -0.0363      0.2301   -0.16      0.87
## Year2010          -0.1282      0.2285   -0.56      0.58
## Year2011          -0.0446      0.2238   -0.20      0.84
## Year2012          -0.0437      0.2357   -0.19      0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0292, Adjusted R-squared:  -0.00111
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 88 is an outlier with |weight| = 0 ( < 0.00018);
## 42 weights are ~= 1. The remaining 519 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.124  0.857  0.950  0.895  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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.149 1          1.072
## Year              1.149 16          1.004

```

Residuals from last author



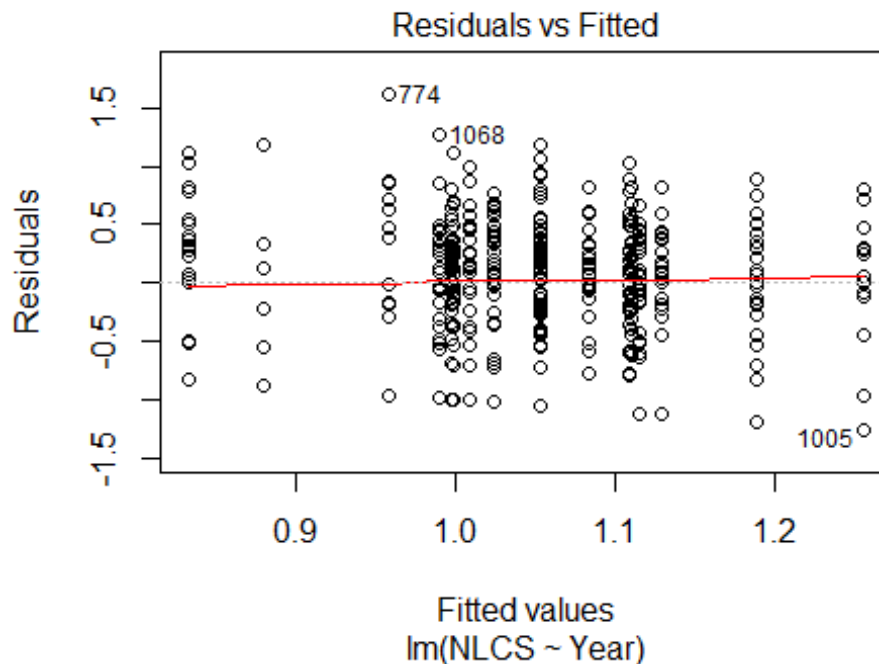
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2780 -0.3189 0.0164 0.3275 2.2968
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22005 0.21576 5.65 2.5e-08 ***
## LastAuthorFemale1 0.00171 0.05578 0.03 0.98
## Year1997 0.04997 0.23892 0.21 0.83
## Year1998 0.05812 0.23475 0.25 0.80
## Year1999 -0.11181 0.24759 -0.45 0.65
## Year2000 -0.04901 0.22858 -0.21 0.83
## Year2001 -0.03492 0.28037 -0.12 0.90
## Year2002 0.03566 0.23968 0.15 0.88
## Year2003 -0.02140 0.23671 -0.09 0.93
## Year2004 0.05797 0.22376 0.26 0.80
## Year2005 -0.06973 0.22480 -0.31 0.76
## Year2006 -0.31171 0.23196 -1.34 0.18
```

```

## Year2007      -0.12454      0.23575      -0.53      0.60
## Year2008      -0.06328      0.23538      -0.27      0.79
## Year2009      -0.03735      0.22982      -0.16      0.87
## Year2010      -0.12783      0.22788      -0.56      0.58
## Year2011      -0.04490      0.22486      -0.20      0.84
## Year2012      -0.04346      0.23512      -0.18      0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0287, Adjusted R-squared:  -0.00162
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 88 is an outlier with |weight| = 0 ( < 0.00018);
## 40 weights are ~= 1. The remaining 521 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.123  0.856  0.949  0.896  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      1.78e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 562"
## [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
##   84  102  96  107  109   76   92  48   76   55   62   81   70   69   82
## 2011 2012
##   76   71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    6   23   23   26   33   21   30   13   23   21   20   26   33   26   33

```

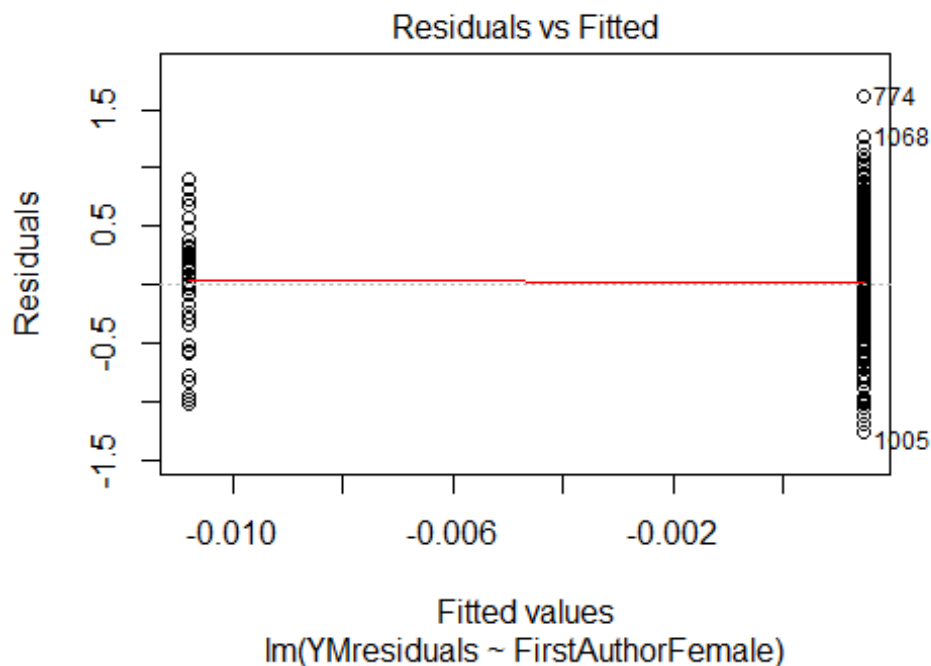
```
## 2011 2012
## 44 41
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 20 21 24 30 19 27 13 22 19 18 23 26 21 29
## 2011 2012
## 37 40
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## 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 = 0.27, df = 1, p-value = 0.6
## [1] "Female first author team size 2018 geometric mean: 2"
## [1] "Male first author team size 2018 geometric mean: 2.126985641715"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

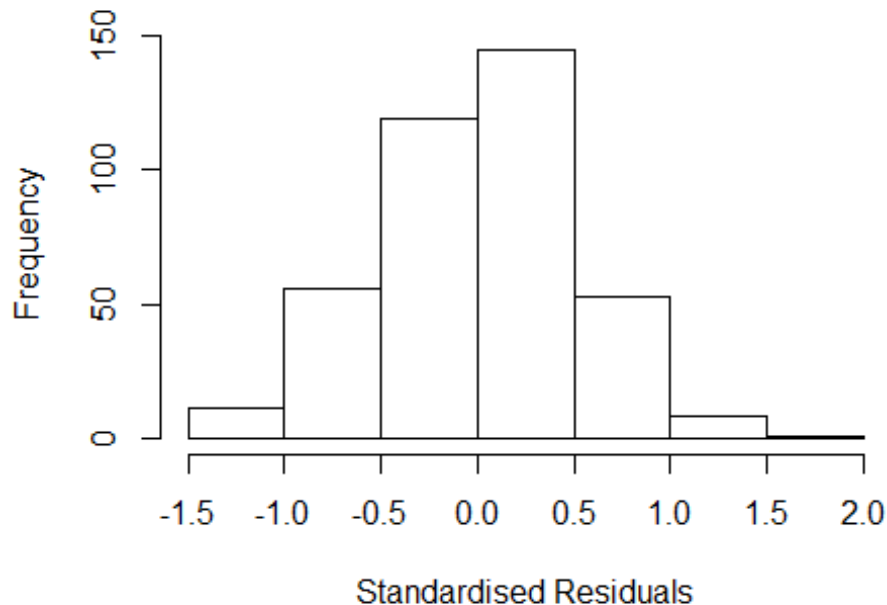
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 8"
## [1] "Male last author team size 2018 geometric mean: 1.96454972488877"

## 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 = 17, 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 1.248 1      1.117
## LastAuthorFemale  1.231 1      1.110
## UniqueAuthors    2.279 4      1.108
## Year              3.012 16     1.035
```

Residuals from first and last author and team size



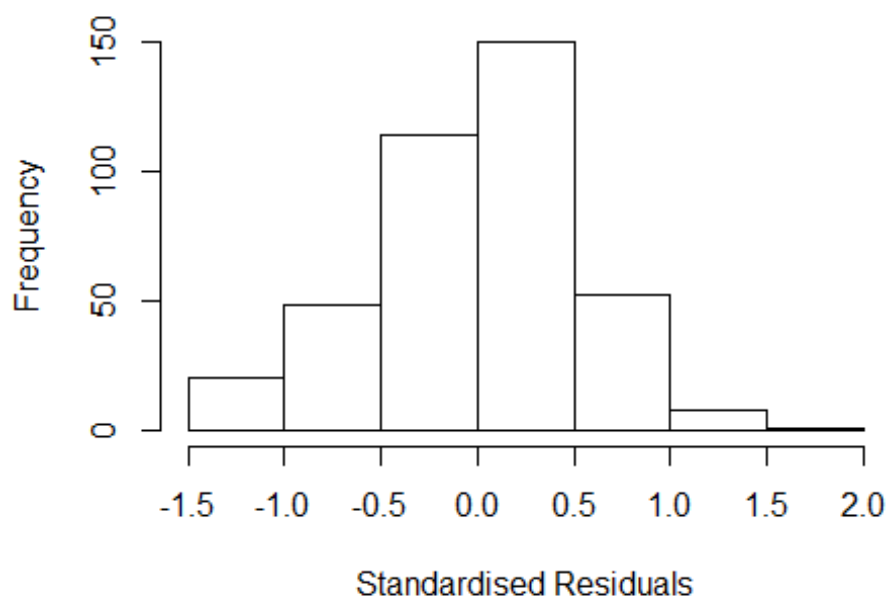
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3530 -0.3724 0.0253 0.3703 1.7742
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90898 0.41593 2.19 0.029 *
## FirstAuthorFemale1 -0.00164 0.08721 -0.02 0.985
## LastAuthorFemale1 0.06176 0.09771 0.63 0.528
## UniqueAuthors2 0.08656 0.08259 1.05 0.295
## UniqueAuthors3 0.19762 0.08361 2.36 0.019 *
## UniqueAuthors4 -0.04353 0.12321 -0.35 0.724
## UniqueAuthors5 0.23449 0.09287 2.52 0.012 *
## Year1997 0.25606 0.43339 0.59 0.555
## Year1998 0.09193 0.43077 0.21 0.831
## Year1999 0.14341 0.43297 0.33 0.741
```

```

## Year2000      -0.03687    0.43802   -0.08    0.933
## Year2001      -0.11123    0.47884   -0.23    0.816
## Year2002      -0.17601    0.44522   -0.40    0.693
## Year2003       0.35743    0.44738    0.80    0.425
## Year2004      -0.05072    0.43153   -0.12    0.906
## Year2005       0.08841    0.42892    0.21    0.837
## Year2006       0.03197    0.44155    0.07    0.942
## Year2007       0.08150    0.43950    0.19    0.853
## Year2008       0.06920    0.42494    0.16    0.871
## Year2009       0.02603    0.42997    0.06    0.952
## Year2010       0.06573    0.42756    0.15    0.878
## Year2011       0.10998    0.43101    0.26    0.799
## Year2012       0.00333    0.42461    0.01    0.994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.0639, Adjusted R-squared:  0.00822
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 363 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  0.876  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      2.54e-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 1.219 1      1.104
## LastAuthorFemale  1.205 1      1.098
## Year              1.462 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
## -1.3108 -0.3797 0.0506 0.3465 1.6827
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94675 0.52184 1.81 0.07 .
## FirstAuthorFemale1 -0.00277 0.08520 -0.03 0.97
## LastAuthorFemale1 0.05767 0.09819 0.59 0.56
## Year1997 0.27068 0.53731 0.50 0.61
## Year1998 0.11218 0.53194 0.21 0.83
## Year1999 0.18991 0.53449 0.36 0.72
## Year2000 0.02973 0.53738 0.06 0.96
## Year2001 -0.05749 0.57935 -0.10 0.92
## Year2002 -0.13977 0.54545 -0.26 0.80
## Year2003 0.36405 0.54537 0.67 0.50
## Year2004 -0.00877 0.53375 -0.02 0.99
## Year2005 0.14163 0.53175 0.27 0.79
```

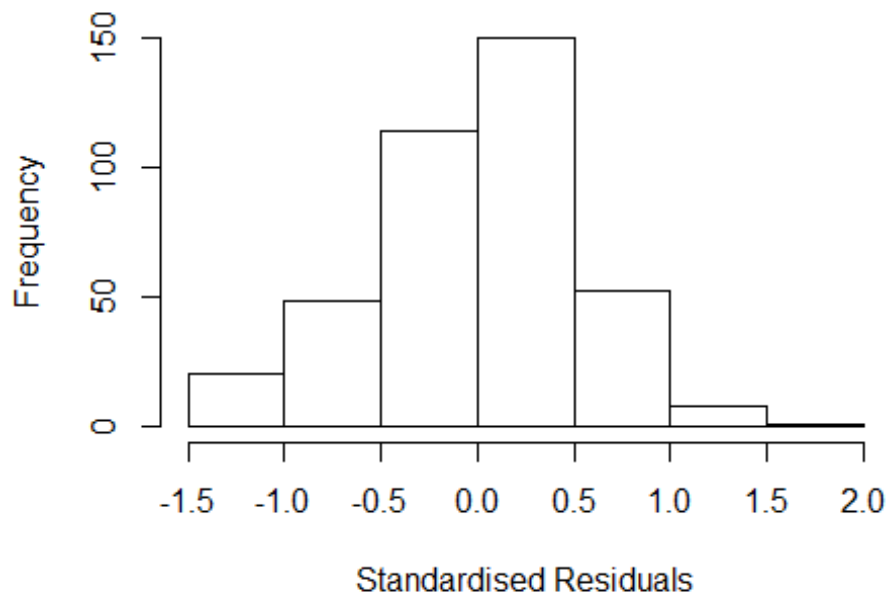


```

## Year2006          0.04573      0.54250      0.08      0.93
## Year2007          0.11738      0.54093      0.22      0.83
## Year2008          0.12935      0.52755      0.25      0.81
## Year2009          0.12344      0.52981      0.23      0.82
## Year2010          0.12876      0.52826      0.24      0.81
## Year2011          0.17222      0.53237      0.32      0.75
## Year2012          0.05608      0.52650      0.11      0.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0384, Adjusted R-squared:  -0.00789
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.273  0.858  0.947  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      2.54e-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 1.201 1      1.096
## Year              1.201 16      1.006

```

Residuals from first author



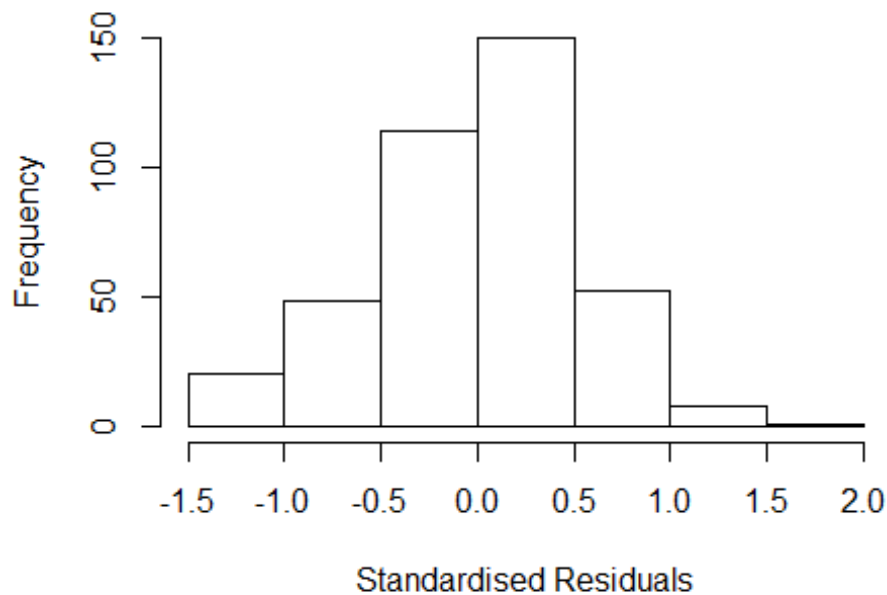
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3131 -0.3816 0.0538 0.3499 1.6858
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94719 0.51961 1.82 0.069 .
## FirstAuthorFemale1 0.00895 0.08613 0.10 0.917
## Year1997 0.27351 0.53452 0.51 0.609
## Year1998 0.11992 0.52950 0.23 0.821
## Year1999 0.19067 0.53238 0.36 0.720
## Year2000 0.04003 0.53463 0.07 0.940
## Year2001 -0.06097 0.57650 -0.11 0.916
## Year2002 -0.13797 0.54339 -0.25 0.800
## Year2003 0.36595 0.54263 0.67 0.500
## Year2004 -0.00549 0.53164 -0.01 0.992
## Year2005 0.15315 0.52869 0.29 0.772
## Year2006 0.04692 0.54042 0.09 0.931
```

```

## Year2007          0.11531      0.53828      0.21      0.830
## Year2008          0.12775      0.52522      0.24      0.808
## Year2009          0.12710      0.52760      0.24      0.810
## Year2010          0.13184      0.52579      0.25      0.802
## Year2011          0.17381      0.52959      0.33      0.743
## Year2012          0.05643      0.52409      0.11      0.914
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0377, Adjusted R-squared:  -0.00595
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 361 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.275  0.869  0.946  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      2.54e-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 1.206 1          1.098
## Year              1.206 16          1.006

```

Residuals from last author



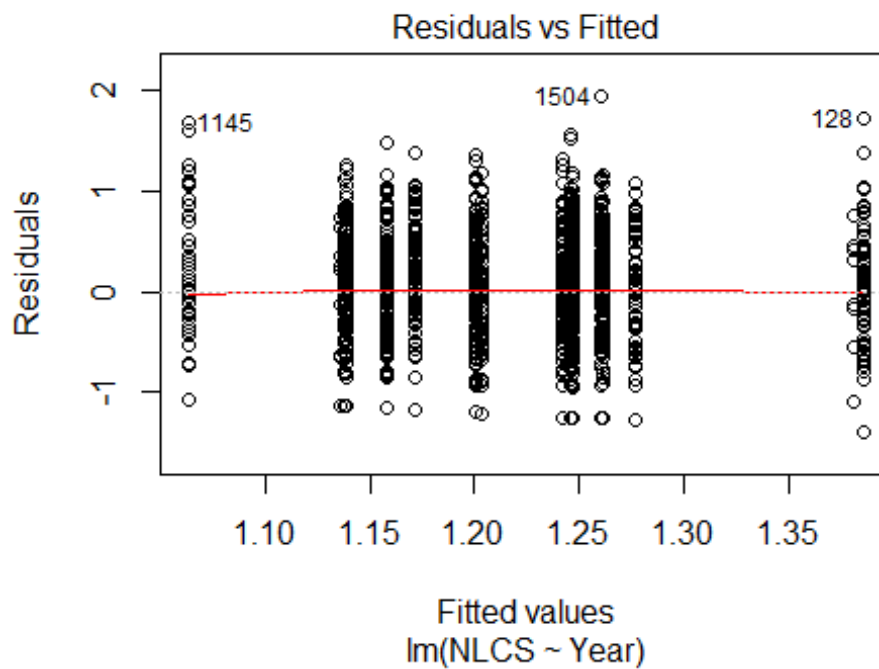
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3107 -0.3792 0.0506 0.3465 1.6842
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94635 0.52324 1.81 0.071 .
## LastAuthorFemale1 0.05696 0.10006 0.57 0.570
## Year1997 0.27129 0.53868 0.50 0.615
## Year1998 0.11260 0.53332 0.21 0.833
## Year1999 0.18990 0.53571 0.35 0.723
## Year2000 0.03015 0.53878 0.06 0.955
## Year2001 -0.05850 0.57729 -0.10 0.919
## Year2002 -0.14009 0.54608 -0.26 0.798
## Year2003 0.36436 0.54689 0.67 0.506
## Year2004 -0.00891 0.53483 -0.02 0.987
## Year2005 0.14208 0.53305 0.27 0.790
## Year2006 0.04564 0.54324 0.08 0.933
```

```

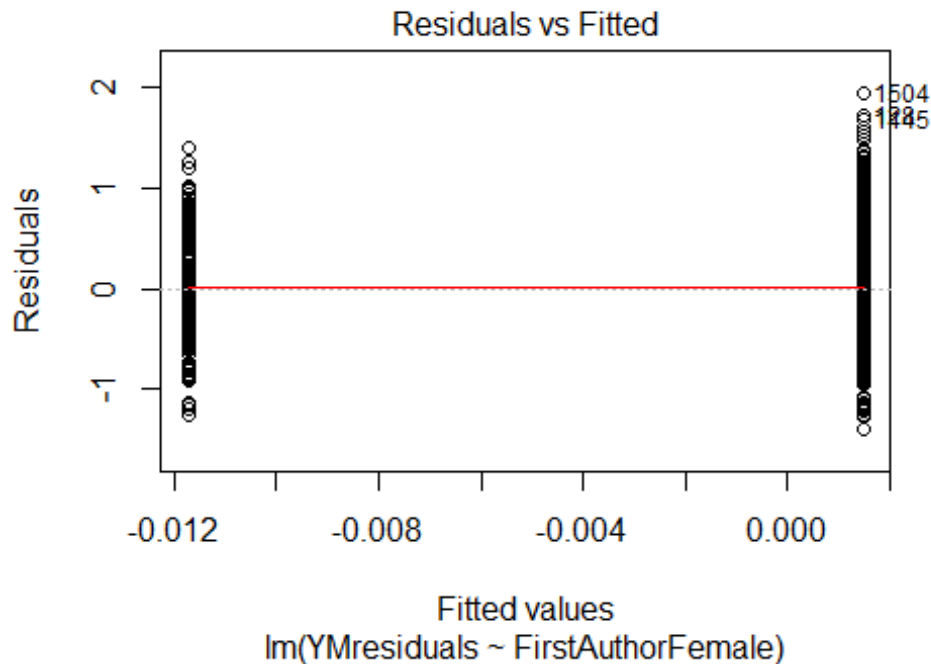
## Year2007      0.11783    0.54222    0.22    0.828
## Year2008      0.12960    0.52886    0.25    0.807
## Year2009      0.12331    0.53092    0.23    0.816
## Year2010      0.12912    0.52960    0.24    0.808
## Year2011      0.17282    0.53378    0.32    0.746
## Year2012      0.05647    0.52796    0.11    0.915
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.0385, Adjusted R-squared:  -0.00504
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 356 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.269  0.858  0.947  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      2.54e-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: 393"
## [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
## 226 168 201 16 196 221 277 275 21 284 22 302 299 340 321
## 2011 2012
## 342 358
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 63 67 7 66 57 95 102 10 125 8 149 133 161 152
## 2011 2012

```

```
## 166 186
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 63 51 48 6 48 47 66 71 8 99 8 123 103 126 124
## 2011 2012
## 129 137
## [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
```

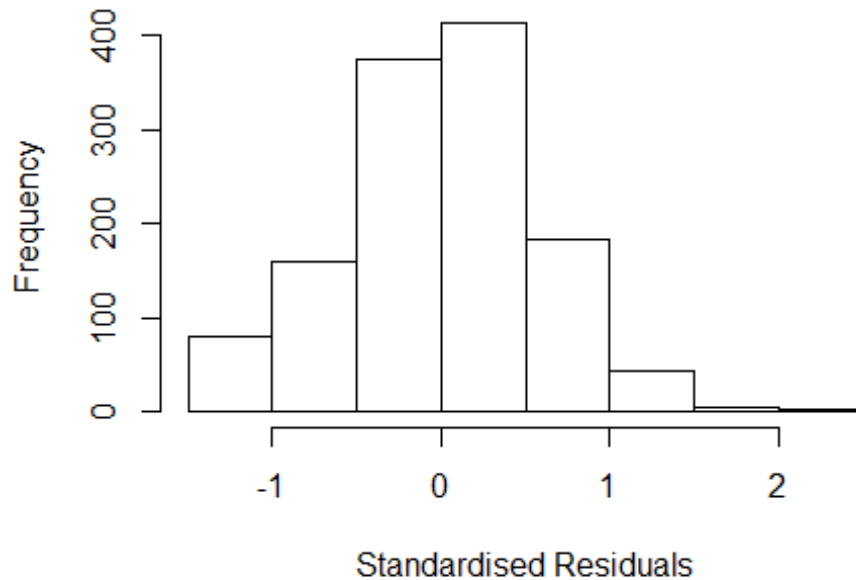


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.91, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.24181594208484"
## [1] "Male first author team size 2018 geometric mean: 3.50453043865283"
##
## 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: 3.59699686349651"
## [1] "Male last author team size 2018 geometric mean: 3.63918862393989"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 400, 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 1.057 1          1.028
## LastAuthorFemale  1.123 1          1.060
## UniqueAuthors     1.425 4          1.045
## 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
## -1.4313 -0.3735 0.0141 0.3639 2.0731
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27542 0.08480 15.04 < 2e-16 ***
## FirstAuthorFemale1 -0.02557 0.05072 -0.50 0.6143
## LastAuthorFemale1 0.00814 0.04758 0.17 0.8642
## UniqueAuthors2 0.03636 0.06011 0.60 0.5453
## UniqueAuthors3 0.09508 0.06041 1.57 0.1158
## UniqueAuthors4 0.17689 0.06342 2.79 0.0054 **
## UniqueAuthors5 0.31913 0.06149 5.19 2.5e-07 ***
## Year1997 -0.03545 0.11261 -0.31 0.7530
## Year1998 -0.25837 0.10065 -2.57 0.0104 *
## Year1999 -0.34243 0.16475 -2.08 0.0379 *
```

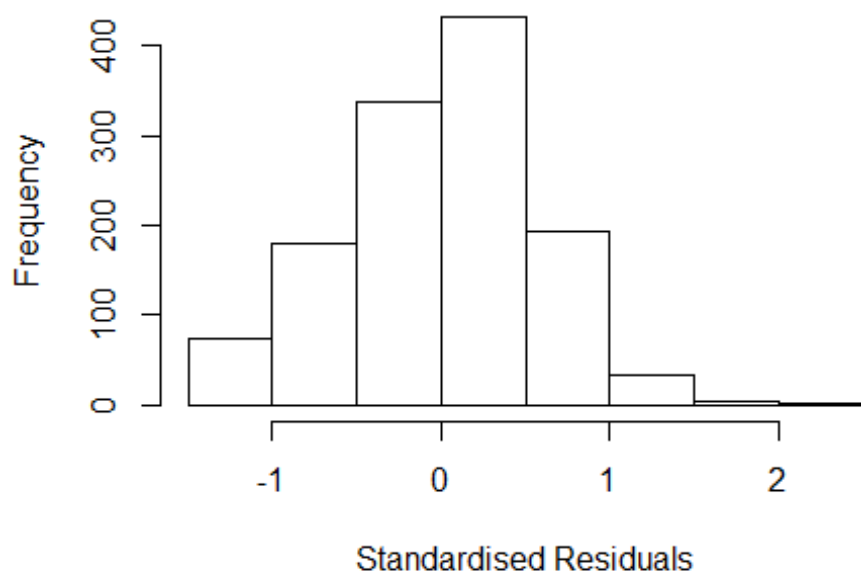


```

## Year2000      -0.21496    0.12443   -1.73    0.0843 .
## Year2001      -0.31046    0.15789   -1.97    0.0495 *
## Year2002      -0.18590    0.11339   -1.64    0.1014 .
## Year2003      -0.18977    0.10766   -1.76    0.0782 .
## Year2004      -0.29660    0.24676   -1.20    0.2296 .
## Year2005      -0.14846    0.09255   -1.60    0.1089 .
## Year2006      -0.01562    0.24952   -0.06    0.9501 .
## Year2007      -0.16497    0.09351   -1.76    0.0779 .
## Year2008      -0.15835    0.09258   -1.71    0.0874 .
## Year2009      -0.17533    0.09225   -1.90    0.0576 .
## Year2010      -0.25336    0.08993   -2.82    0.0049 **
## Year2011      -0.17140    0.09227   -1.86    0.0635 .
## Year2012      -0.25896    0.09397   -2.76    0.0059 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0485, Adjusted R-squared:  0.0315
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1135 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.133  0.865  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      7.96e-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 1.042 1      1.021
## LastAuthorFemale  1.072 1      1.035
## Year              1.114 16      1.003

```

Residuals from first and last author



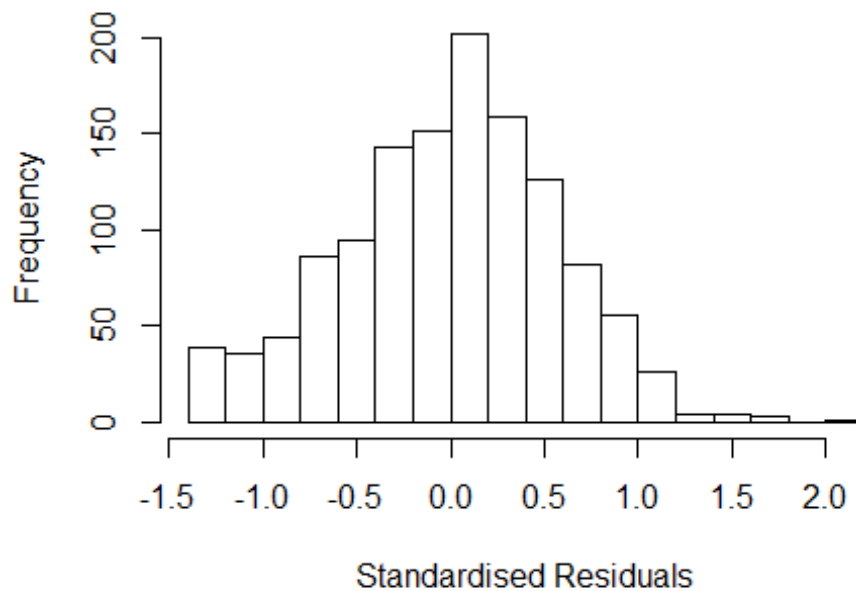
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.346 -0.368 0.030 0.379 2.016
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.34571 0.07640 17.61 <2e-16 ***
## FirstAuthorFemale1 0.00994 0.05279 0.19 0.8506
## LastAuthorFemale1 0.03231 0.04903 0.66 0.5101
## Year1997 -0.05096 0.11074 -0.46 0.6455
## Year1998 -0.26647 0.10291 -2.59 0.0097 **
## Year1999 -0.26656 0.20378 -1.31 0.1911
## Year2000 -0.19061 0.12460 -1.53 0.1263
## Year2001 -0.32075 0.15481 -2.07 0.0385 *
## Year2002 -0.16236 0.11336 -1.43 0.1523
## Year2003 -0.14296 0.10865 -1.32 0.1885
## Year2004 -0.31855 0.24810 -1.28 0.1994
## Year2005 -0.10175 0.09230 -1.10 0.2705
```

```

## Year2006          0.07146      0.22787      0.31      0.7539
## Year2007          -0.11345      0.09347     -1.21      0.2251
## Year2008          -0.09735      0.09346     -1.04      0.2978
## Year2009          -0.11543      0.09360     -1.23      0.2177
## Year2010          -0.19252      0.08993     -2.14      0.0325 *
## Year2011          -0.12125      0.09154     -1.32      0.1856
## Year2012          -0.16507      0.09224     -1.79      0.0737 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.0138, Adjusted R-squared:  -0.000586
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1143 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.184  0.868  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      7.96e-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 1.04 1      1.020
## Year              1.04 16      1.001

```

Residuals from first author



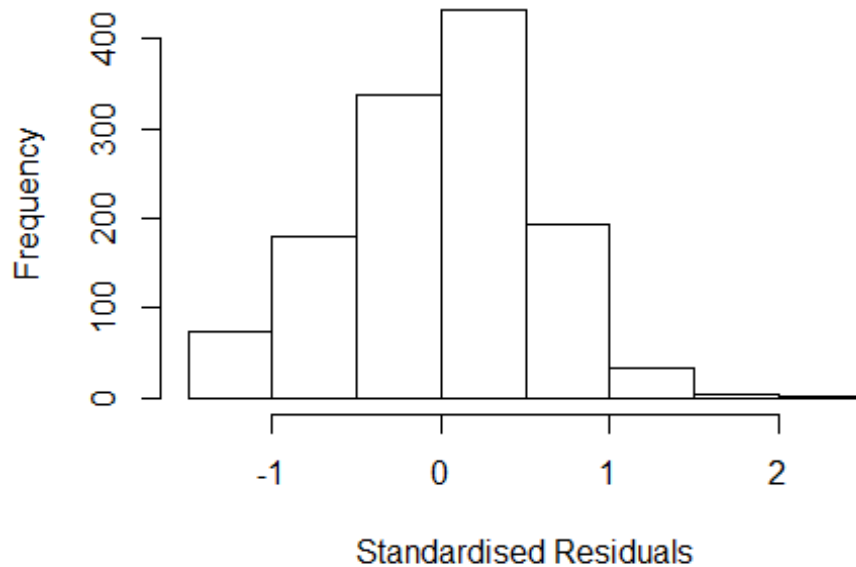
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3464 -0.3729  0.0278  0.3838  2.0118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3464    0.0764   17.63  <2e-16 ***
## FirstAuthorFemale1  0.0111    0.0530    0.21   0.834
## Year1997        -0.0478    0.1107   -0.43   0.666
## Year1998        -0.2650    0.1029   -2.58   0.010 *
## Year1999        -0.2672    0.2037   -1.31   0.190
## Year2000        -0.1858    0.1238   -1.50   0.133
## Year2001        -0.3160    0.1541   -2.05   0.041 *
## Year2002        -0.1592    0.1133   -1.41   0.160
## Year2003        -0.1412    0.1088   -1.30   0.195
## Year2004        -0.3147    0.2500   -1.26   0.208
## Year2005        -0.0989    0.0923   -1.07   0.284
## Year2006         0.0706    0.2278    0.31   0.757
```

```

## Year2007          -0.1113      0.0936   -1.19    0.235
## Year2008          -0.0935      0.0935   -1.00    0.317
## Year2009          -0.1109      0.0934   -1.19    0.235
## Year2010          -0.1902      0.0899   -2.12    0.035 *
## Year2011          -0.1140      0.0907   -1.26    0.209
## Year2012          -0.1629      0.0921   -1.77    0.077 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.0134, Adjusted R-squared:  -0.000138
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1140 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.188  0.868  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      7.96e-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 1.069 1      1.034
## Year            1.069 16      1.002

```

Residuals from last author



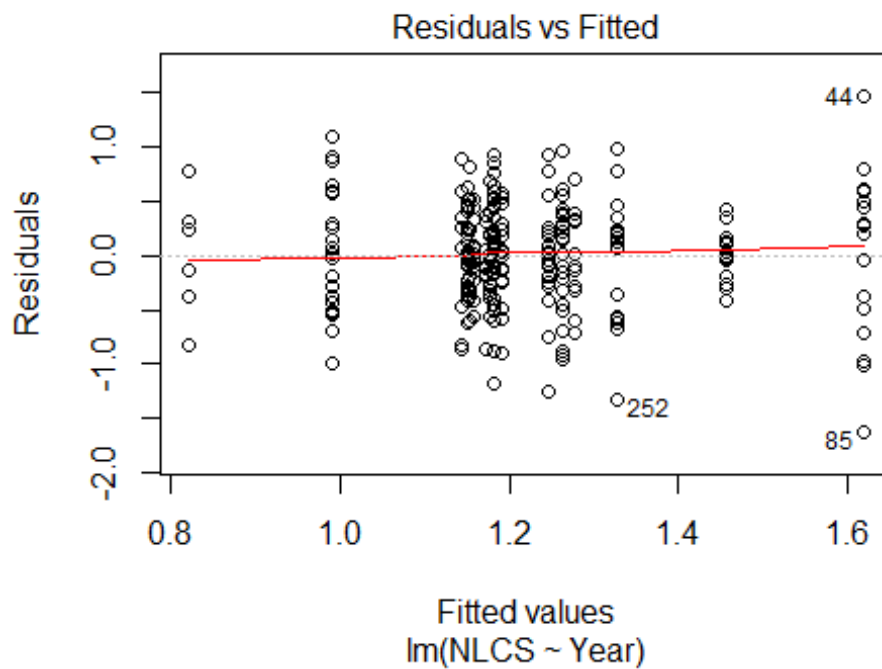
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3463 -0.3691 0.0295 0.3787 2.0143
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3463 0.0762 17.67 <2e-16 ***
## LastAuthorFemale1 0.0327 0.0491 0.66 0.5062
## Year1997 -0.0516 0.1106 -0.47 0.6412
## Year1998 -0.2656 0.1029 -2.58 0.0099 **
## Year1999 -0.2671 0.2037 -1.31 0.1900
## Year2000 -0.1901 0.1250 -1.52 0.1286
## Year2001 -0.3208 0.1549 -2.07 0.0386 *
## Year2002 -0.1616 0.1133 -1.43 0.1540
## Year2003 -0.1432 0.1086 -1.32 0.1874
## Year2004 -0.3191 0.2481 -1.29 0.1985
## Year2005 -0.1011 0.0923 -1.10 0.2735
## Year2006 0.0710 0.2278 0.31 0.7555
```

```

## Year2007          -0.1128      0.0935   -1.21    0.2279
## Year2008          -0.0972      0.0935   -1.04    0.2989
## Year2009          -0.1147      0.0936   -1.23    0.2207
## Year2010          -0.1913      0.0899   -2.13    0.0337 *
## Year2011          -0.1204      0.0916   -1.31    0.1890
## Year2012          -0.1647      0.0923   -1.78    0.0747 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0137, Adjusted R-squared:  0.000201
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1141 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.184  0.868  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      7.96e-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: 1257"
## [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
##   34   44   34   25   31   47   32   25   33   16   41   25   28   29   44
## 2011 2012
##   30   50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   16   19   14   12   18   16   13   19    6   23   15    9   13   25
## 2011 2012

```

```
## 22 27
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 15 18 12 12 15 16 13 17 5 17 14 9 13 21
## 2011 2012
## 19 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 = 41, df = 16, p-value = 5e-04
```

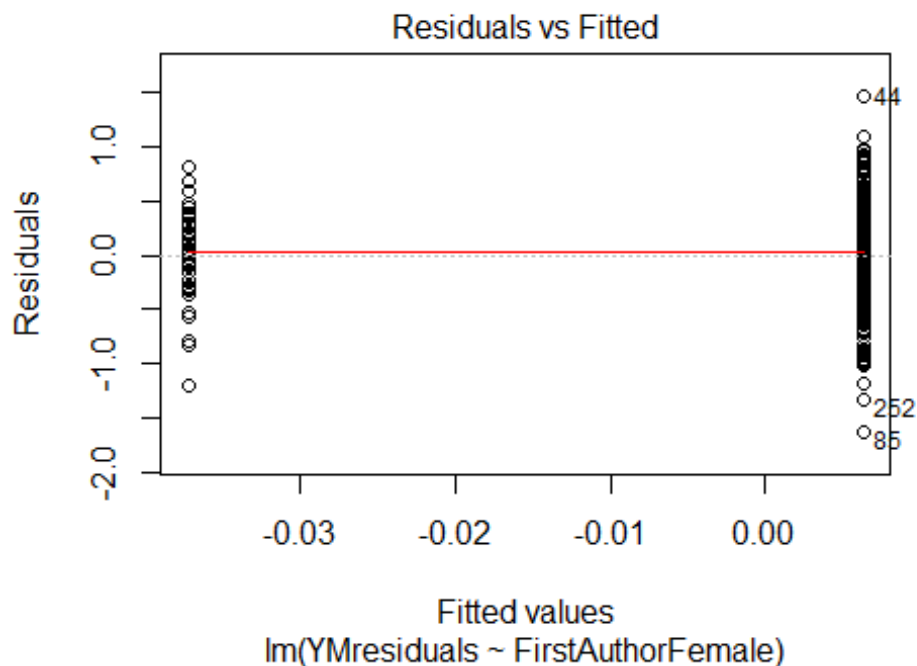


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 5.27726567472935"
## [1] "Male first author team size 2018 geometric mean: 3.99841972589167"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



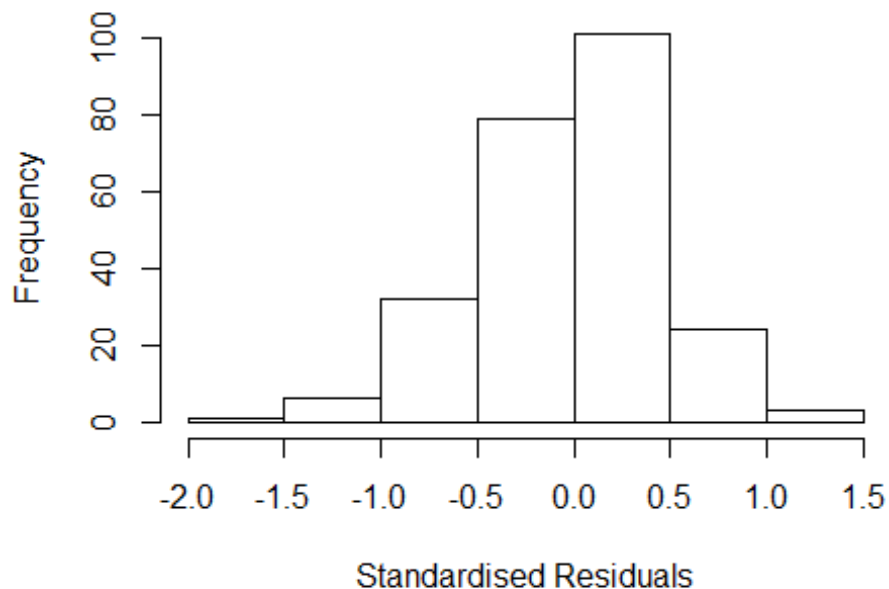
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 88, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.22419991126449"
## [1] "Male last author team size 2018 geometric mean: 4.24982262942477"

## 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 = 77, 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 1.556 1          1.248
## LastAuthorFemale  1.424 1          1.194
## UniqueAuthors    3.104 4          1.152
## Year              5.545 16         1.055
```

Residuals from first and last author and team size



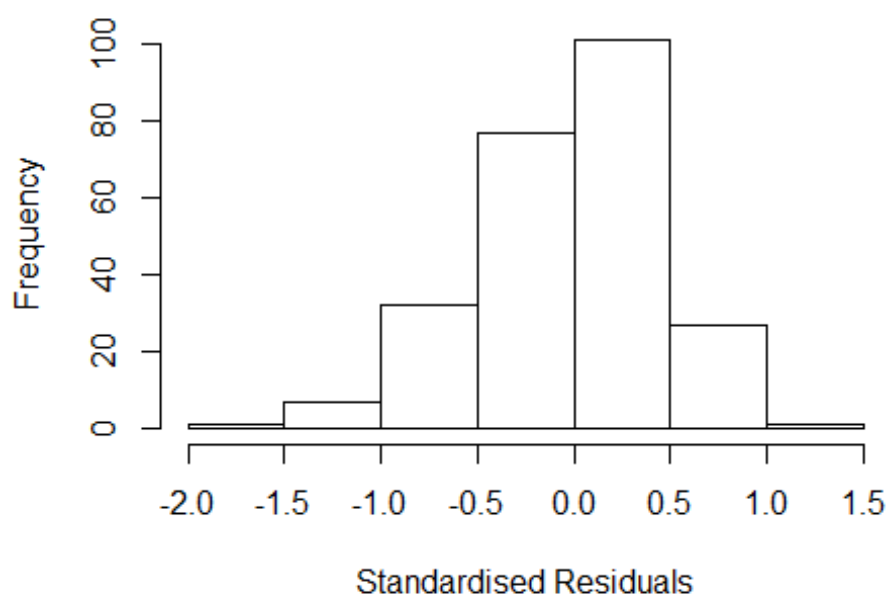
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8489 -0.2341 0.0133 0.2489 1.2361
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17969 0.42479 2.78 0.006 **
## FirstAuthorFemale1 -0.10100 0.08613 -1.17 0.242
## LastAuthorFemale1 0.02368 0.07740 0.31 0.760
## UniqueAuthors2 -0.29002 0.14600 -1.99 0.048 *
## UniqueAuthors3 -0.10694 0.13096 -0.82 0.415
## UniqueAuthors4 -0.04780 0.13001 -0.37 0.713
## UniqueAuthors5 0.12960 0.11351 1.14 0.255
## Year1997 0.66917 0.48848 1.37 0.172
## Year1998 0.05824 0.42860 0.14 0.892
## Year1999 0.01320 0.41898 0.03 0.975
```

```

## Year2000      0.09796      0.42388      0.23      0.817
## Year2001      0.23806      0.43486      0.55      0.585
## Year2002      0.02707      0.42526      0.06      0.949
## Year2003      0.09889      0.43193      0.23      0.819
## Year2004     -0.04608      0.42310     -0.11      0.913
## Year2005     -0.13803      0.53071     -0.26      0.795
## Year2006     -0.17791      0.44471     -0.40      0.689
## Year2007      0.06269      0.45260      0.14      0.890
## Year2008     -0.00669      0.43897     -0.02      0.988
## Year2009      0.27764      0.41935      0.66      0.509
## Year2010     -0.02960      0.41800     -0.07      0.944
## Year2011      0.09817      0.42832      0.23      0.819
## Year2012      0.05148      0.42953      0.12      0.905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.207, Adjusted R-squared:  0.129
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 225 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0142 0.8490 0.9630 0.8900 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      4.07e-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 1.466 1      1.211
## LastAuthorFemale 1.228 1      1.108
## Year      1.781 16      1.018

```

Residuals from first and last author



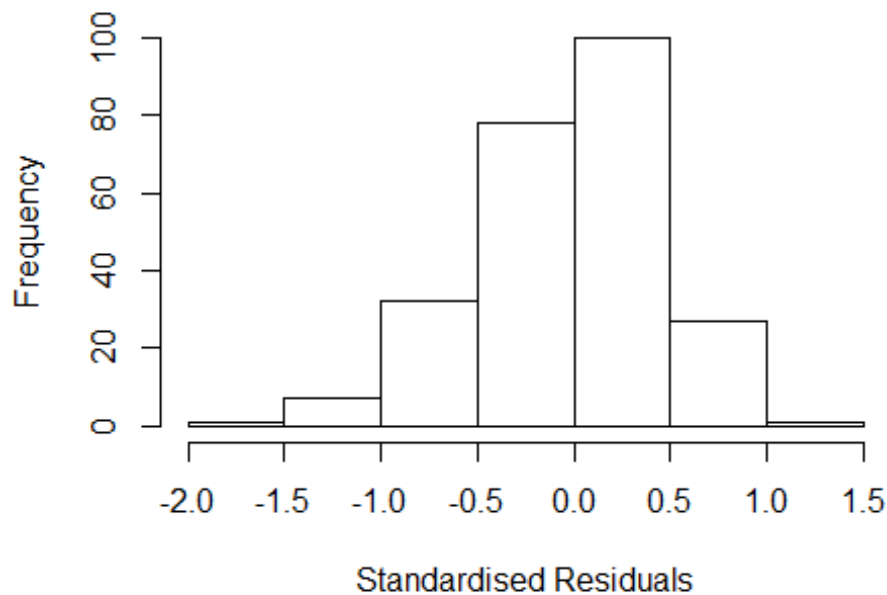
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.765 -0.288 0.017 0.298 1.320
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10624 0.36578 3.02 0.0028 **
## FirstAuthorFemale1 -0.06632 0.08128 -0.82 0.4154
## LastAuthorFemale1 0.00914 0.07756 0.12 0.9063
## Year1997 0.65922 0.46962 1.40 0.1618
## Year1998 0.12944 0.38030 0.34 0.7339
## Year1999 0.06995 0.36953 0.19 0.8500
## Year2000 0.08537 0.37759 0.23 0.8213
## Year2001 0.33085 0.39698 0.83 0.4055
## Year2002 0.07755 0.37475 0.21 0.8362
## Year2003 0.18040 0.38402 0.47 0.6390
## Year2004 0.05204 0.37619 0.14 0.8901
## Year2005 -0.13976 0.45403 -0.31 0.7585
```

```

## Year2006      -0.16053    0.40031   -0.40    0.6888
## Year2007      0.14538    0.40833    0.36    0.7221
## Year2008      0.03988    0.39457    0.10    0.9196
## Year2009      0.36711    0.36916    0.99    0.3211
## Year2010      0.04451    0.37232    0.12    0.9049
## Year2011      0.17296    0.38714    0.45    0.6555
## Year2012      0.15338    0.38369    0.40    0.6897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0562
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 215 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0693 0.8340 0.9430 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      4.07e-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 1.482 1      1.218
## Year      1.482 16      1.012

```

Residuals from first author



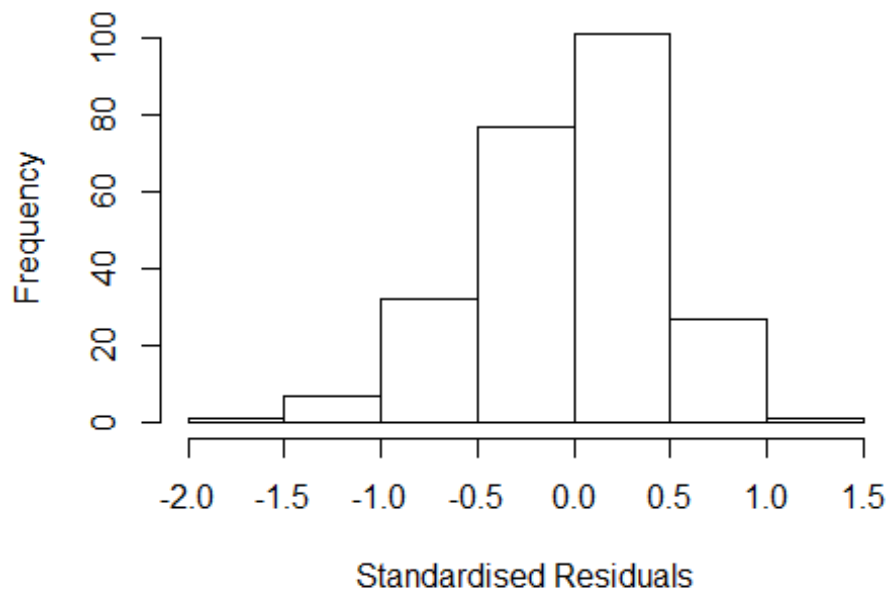
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7686 -0.2909 0.0149 0.2983 1.3164
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1059 0.3678 3.01 0.0029 **
## FirstAuthorFemale1 -0.0646 0.0822 -0.79 0.4323
## Year1997 0.6627 0.4709 1.41 0.1607
## Year1998 0.1309 0.3821 0.34 0.7323
## Year1999 0.0725 0.3706 0.20 0.8450
## Year2000 0.0859 0.3796 0.23 0.8211
## Year2001 0.3313 0.3990 0.83 0.4072
## Year2002 0.0788 0.3761 0.21 0.8342
## Year2003 0.1821 0.3855 0.47 0.6370
## Year2004 0.0541 0.3776 0.14 0.8863
## Year2005 -0.1359 0.4555 -0.30 0.7657
## Year2006 -0.1607 0.4023 -0.40 0.6900
```

```

## Year2007          0.1493      0.4069      0.37      0.7140
## Year2008          0.0421      0.3944      0.11      0.9152
## Year2009          0.3691      0.3702      1.00      0.3198
## Year2010          0.0465      0.3734      0.12      0.9010
## Year2011          0.1758      0.3893      0.45      0.6520
## Year2012          0.1554      0.3850      0.40      0.6869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0613
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 216 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0638 0.8310 0.9430 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      4.07e-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 1.257 1      1.121
## Year              1.257 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
## -1.7670 -0.2903 0.0196 0.2996 1.3180
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0900 0.3534 3.08 0.0023 **
## LastAuthorFemale1 -0.0014 0.0780 -0.02 0.9857
## Year1997 0.6770 0.4523 1.50 0.1358
## Year1998 0.1446 0.3673 0.39 0.6943
## Year1999 0.0623 0.3605 0.17 0.8630
## Year2000 0.1017 0.3648 0.28 0.7808
## Year2001 0.3173 0.3879 0.82 0.4141
## Year2002 0.0814 0.3659 0.22 0.8242
## Year2003 0.1920 0.3736 0.51 0.6079
## Year2004 0.0655 0.3647 0.18 0.8576
## Year2005 -0.1467 0.4439 -0.33 0.7414
## Year2006 -0.1516 0.3891 -0.39 0.6973
```

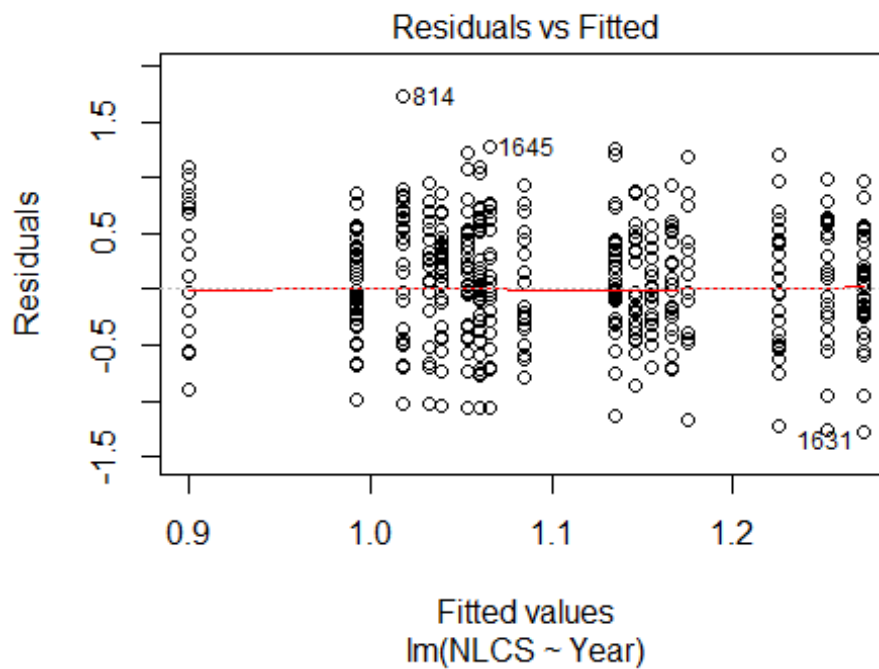


```

## Year2007          0.1543      0.3998      0.39      0.6998
## Year2008          0.0398      0.3865      0.10      0.9181
## Year2009          0.3663      0.3603      1.02      0.3104
## Year2010          0.0561      0.3617      0.15      0.8770
## Year2011          0.1866      0.3715      0.50      0.6159
## Year2012          0.1633      0.3724      0.44      0.6614
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.124, Adjusted R-squared:  0.0582
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0674 0.8320 0.9440 0.8840 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-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 246"
## [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
##   98   72   85   80   57  106   56   99   81   67   98   66   76   84   63
## 2011 2012
##   58   67
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   25   26   21   24   27   16   35   22   21   40   31   35   45   33
## 2011 2012

```

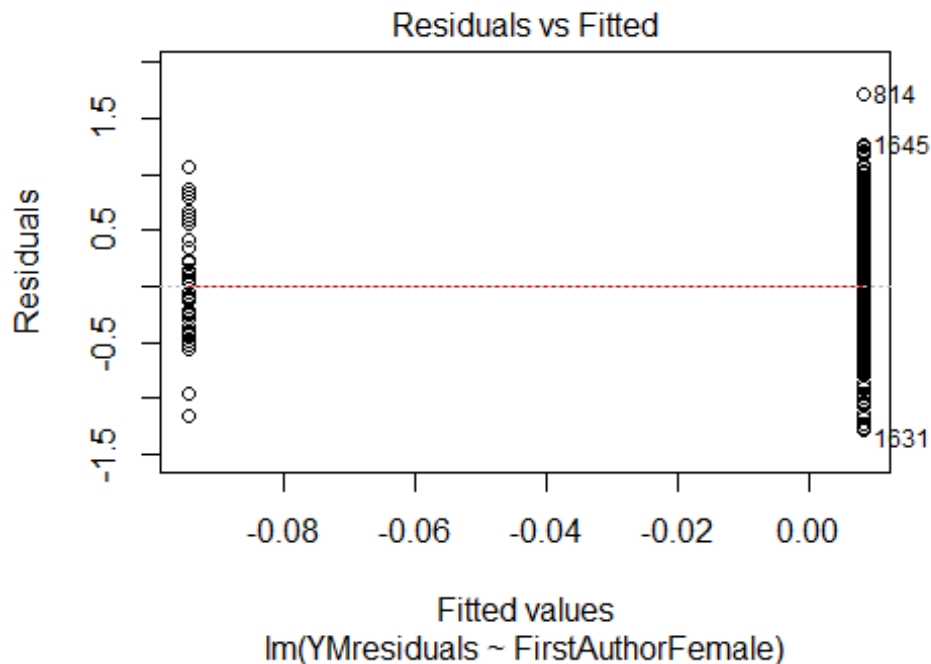
```
## 24 39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 21 15 20 20 26 12 30 17 17 33 29 31 40 29
## 2011 2012
## 21 33
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: 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 = 0.13, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 2.6207413942089"
## [1] "Male first author team size 2018 geometric mean: 2.27653695019436"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

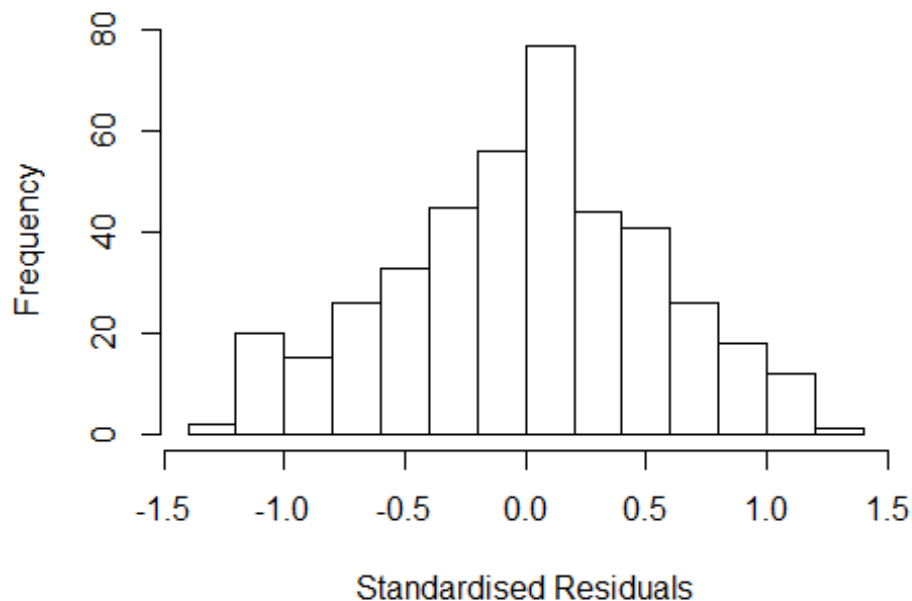
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 56, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.82586236554478"
## [1] "Male last author team size 2018 geometric mean: 2.19720760543476"

## 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 = 0.1
## 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 1.406 1      1.186
## LastAuthorFemale  1.234 1      1.111
## UniqueAuthors    1.894 4      1.083
## Year              2.304 16     1.026
```

Residuals from first and last author and team size



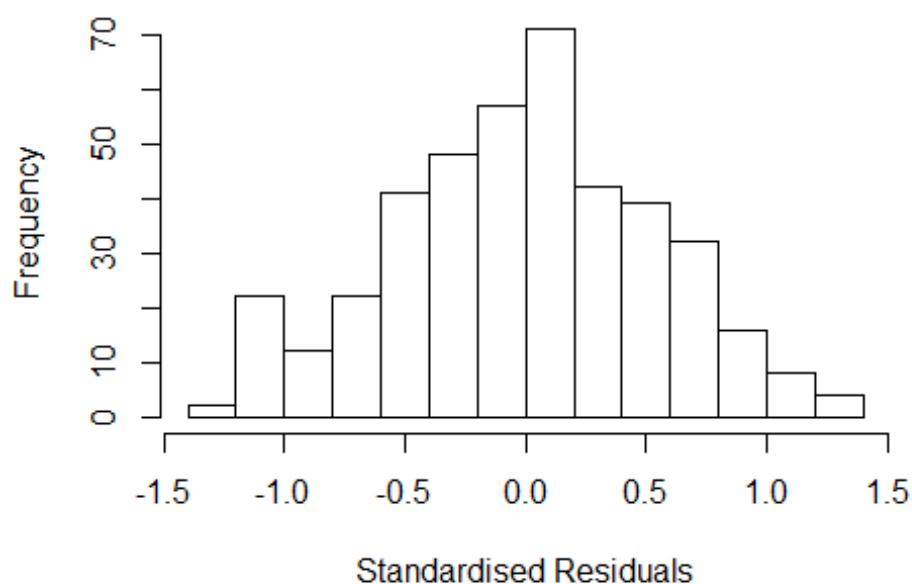
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2948 -0.3728 0.0274 0.3688 1.2583
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1336 0.1404 8.07 8.4e-15 ***
## FirstAuthorFemale1 -0.2070 0.0936 -2.21 0.02758 *
## LastAuthorFemale1 -0.0411 0.1251 -0.33 0.74283
## UniqueAuthors2 0.0433 0.0714 0.61 0.54405
## UniqueAuthors3 0.2807 0.0832 3.37 0.00082 ***
## UniqueAuthors4 0.1001 0.1015 0.99 0.32437
## UniqueAuthors5 0.0722 0.2208 0.33 0.74381
## Year1997 -0.0526 0.1880 -0.28 0.77996
## Year1998 -0.1791 0.1851 -0.97 0.33390
## Year1999 -0.0115 0.1621 -0.07 0.94339
```

```

## Year2000          -0.1339      0.1544   -0.87  0.38634
## Year2001          -0.3334      0.2796   -1.19  0.23371
## Year2002          -0.0781      0.2308   -0.34  0.73506
## Year2003          -0.1196      0.2166   -0.55  0.58119
## Year2004          -0.0877      0.2231   -0.39  0.69463
## Year2005          -0.1616      0.1675   -0.96  0.33532
## Year2006          -0.1578      0.1642   -0.96  0.33726
## Year2007          -0.1508      0.1702   -0.89  0.37591
## Year2008          -0.1160      0.1602   -0.72  0.46974
## Year2009          -0.1718      0.1436   -1.20  0.23206
## Year2010           0.0312      0.1519    0.21  0.83746
## Year2011          -0.1740      0.1846   -0.94  0.34664
## Year2012          -0.1026      0.1531   -0.67  0.50308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.0611, Adjusted R-squared:  0.00856
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 379 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.505  0.852  0.942  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      2.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.289 1      1.135
## LastAuthorFemale  1.174 1      1.083
## Year              1.319 16      1.009

```

Residuals from first and last author



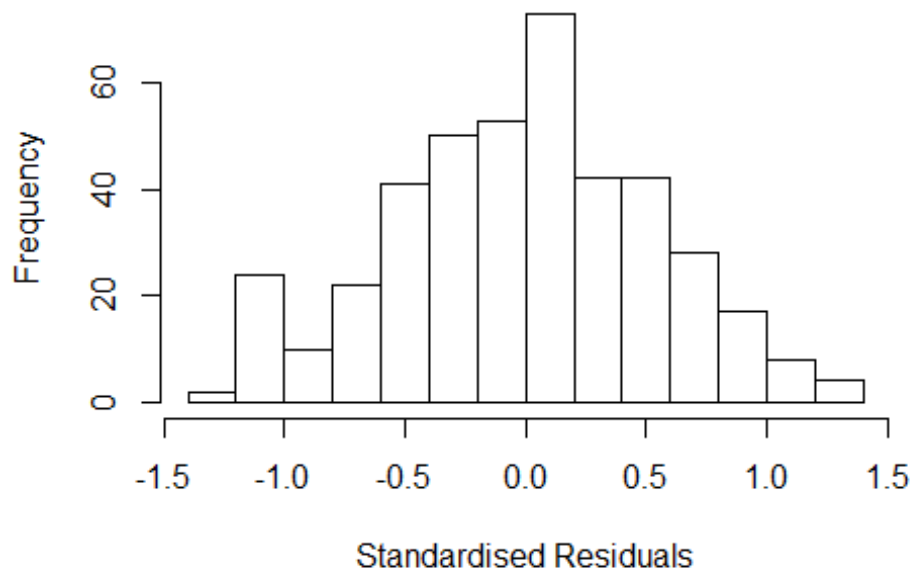
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.244 -0.378 0.016 0.374 1.289
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20930 0.12637 9.57 <2e-16 ***
## FirstAuthorFemale1 -0.16758 0.09241 -1.81 0.071 .
## LastAuthorFemale1 -0.07654 0.12295 -0.62 0.534
## Year1997 -0.06489 0.18191 -0.36 0.721
## Year1998 -0.18978 0.18834 -1.01 0.314
## Year1999 -0.00501 0.16992 -0.03 0.976
## Year2000 -0.10746 0.16039 -0.67 0.503
## Year2001 -0.32549 0.26059 -1.25 0.212
## Year2002 -0.08517 0.22527 -0.38 0.706
## Year2003 -0.12462 0.21628 -0.58 0.565
## Year2004 -0.09779 0.20757 -0.47 0.638
## Year2005 -0.15768 0.17114 -0.92 0.357
```

```

## Year2006      -0.16820    0.16420   -1.02    0.306
## Year2007      -0.14948    0.16988   -0.88    0.379
## Year2008      -0.13922    0.16292   -0.85    0.393
## Year2009      -0.19711    0.14128   -1.40    0.164
## Year2010       0.03499    0.14689    0.24    0.812
## Year2011      -0.13688    0.18798   -0.73    0.467
## Year2012      -0.10077    0.14979   -0.67    0.502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0295, Adjusted R-squared:  -0.0145
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 380 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.519  0.843   0.939   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      2.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.192 1      1.092
## Year              1.192 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
## -1.2413 -0.3683 0.0103 0.3761 1.2987
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20285 0.12565 9.57 <2e-16 ***
## FirstAuthorFemale1 -0.18704 0.08965 -2.09 0.038 *
## Year1997 -0.05990 0.18177 -0.33 0.742
## Year1998 -0.18883 0.18804 -1.00 0.316
## Year1999 -0.00238 0.17059 -0.01 0.989
## Year2000 -0.09994 0.15975 -0.63 0.532
## Year2001 -0.31780 0.26091 -1.22 0.224
## Year2002 -0.08496 0.22658 -0.37 0.708
## Year2003 -0.11758 0.21625 -0.54 0.587
## Year2004 -0.09094 0.20725 -0.44 0.661
## Year2005 -0.15241 0.16879 -0.90 0.367
## Year2006 -0.16898 0.16321 -1.04 0.301
```

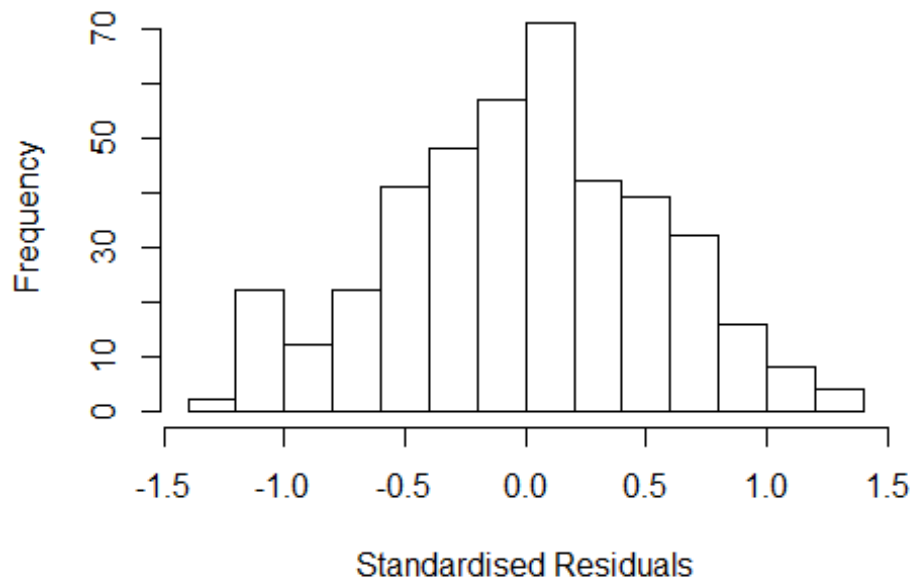


```

## Year2007          -0.15199    0.16846   -0.90    0.367
## Year2008          -0.13482    0.16317   -0.83    0.409
## Year2009          -0.19339    0.14103   -1.37    0.171
## Year2010           0.03842    0.14688    0.26    0.794
## Year2011          -0.12710    0.18672   -0.68    0.496
## Year2012          -0.10351    0.14911   -0.69    0.488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0284, Adjusted R-squared:  -0.0131
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 375 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.510  0.841  0.938  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      2.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.082 1          1.040
## Year            1.082 16          1.002

```

Residuals from last author



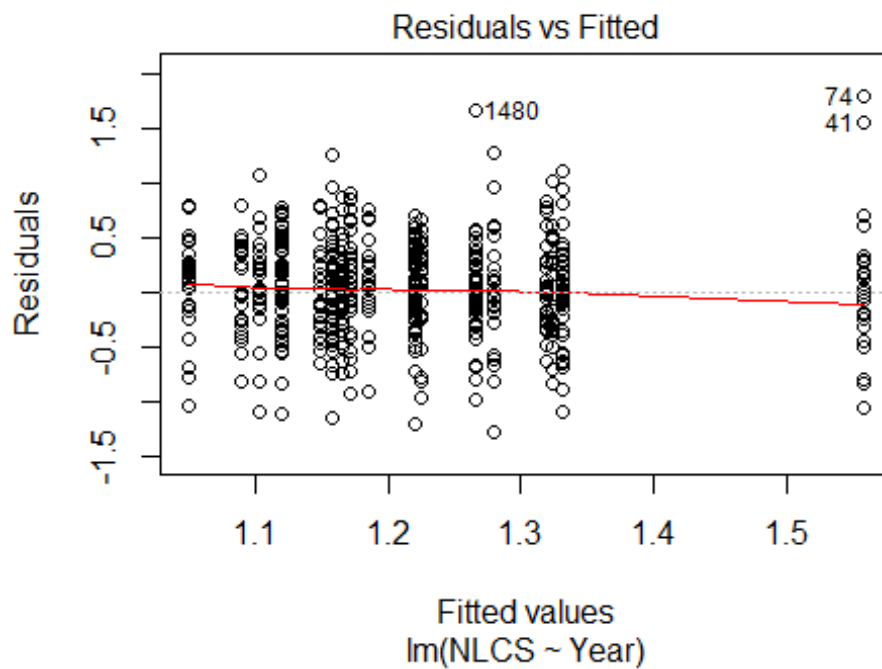
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22197 -0.38694 -0.00428 0.37878 1.30703
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19669 0.12841 9.32 <2e-16 ***
## LastAuthorFemale1 -0.12807 0.11983 -1.07 0.29
## Year1997 -0.07672 0.18558 -0.41 0.68
## Year1998 -0.18344 0.18653 -0.98 0.33
## Year1999 0.00342 0.16847 0.02 0.98
## Year2000 -0.10409 0.16263 -0.64 0.52
## Year2001 -0.32410 0.25364 -1.28 0.20
## Year2002 -0.08776 0.22897 -0.38 0.70
## Year2003 -0.11476 0.21491 -0.53 0.59
## Year2004 -0.08702 0.20787 -0.42 0.68
## Year2005 -0.16338 0.17652 -0.93 0.36
## Year2006 -0.15830 0.16540 -0.96 0.34
```

```

## Year2007          -0.14449      0.17041    -0.85      0.40
## Year2008          -0.14479      0.16349    -0.89      0.38
## Year2009          -0.19513      0.14331    -1.36      0.17
## Year2010           0.02528      0.14840      0.17      0.86
## Year2011          -0.15419      0.18915    -0.82      0.42
## Year2012          -0.08199      0.15212    -0.54      0.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0239, Adjusted R-squared:  -0.0177
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 371 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.522  0.846  0.939  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      2.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 416"
## [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
##   97   83   71   77   76   64   74   77   66   74   99   88  101   94  124
## 2011 2012
##  126   98
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   26   23   19   29   29   25   34   22   21   23   38   32   49   57   52
## 2011 2012

```

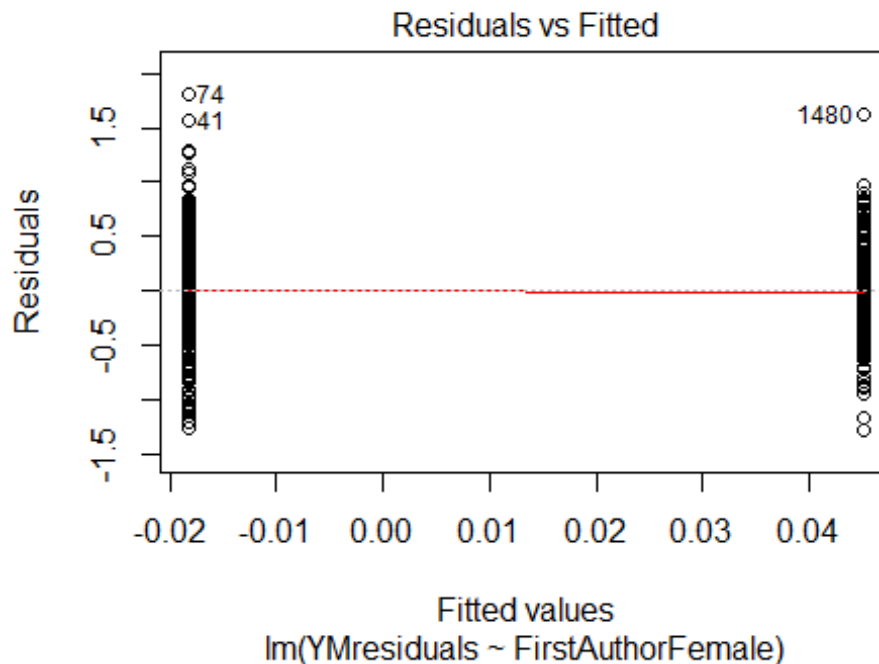
```
## 58 55
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 23 19 18 29 27 22 29 19 19 20 30 29 43 43 46
## 2011 2012
## 48 45
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



```
##
## 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: 4.37951913988789"
## [1] "Male first author team size 2018 geometric mean: 4.85050620748749"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

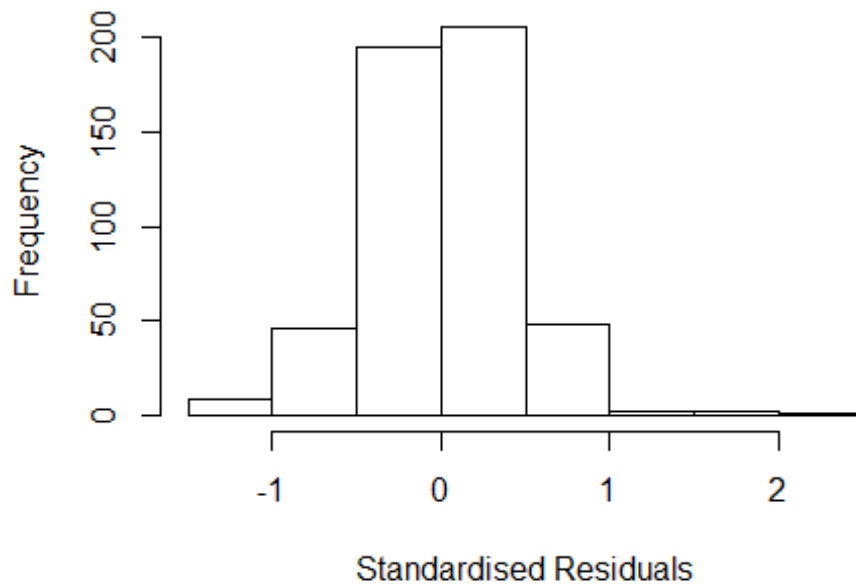
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.73813722053759"
## [1] "Male last author team size 2018 geometric mean: 4.78221890906943"

## 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 = 29, 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 1.179  1      1.086
## LastAuthorFemale  1.326  1      1.151
## UniqueAuthors    2.262  4      1.107
## Year              2.662 16      1.031
```

Residuals from first and last author and team size



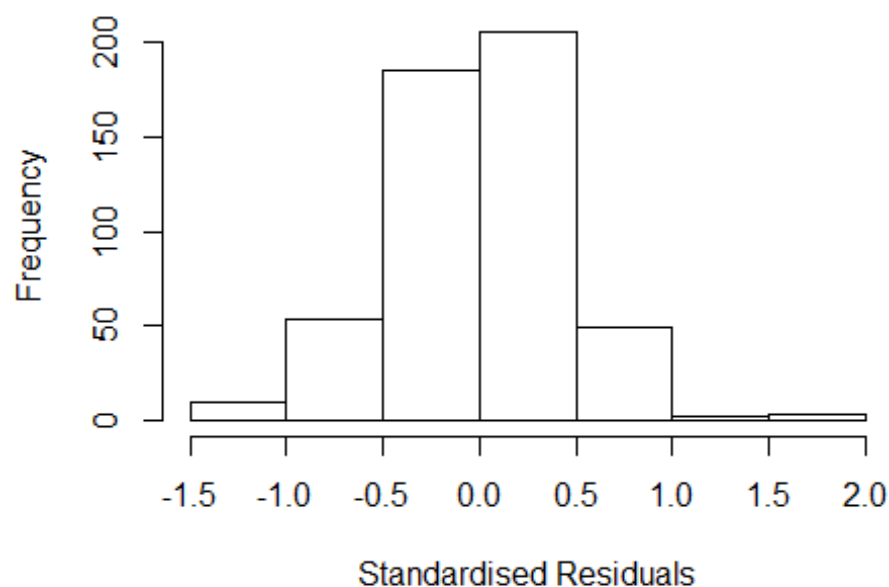
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3256 -0.2491 0.0126 0.2523 2.0725
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2715 0.1584 8.02 7.7e-15 ***
## FirstAuthorFemale1 0.0316 0.0418 0.76 0.4496
## LastAuthorFemale1 -0.0259 0.0671 -0.39 0.6993
## UniqueAuthors2 0.1764 0.1173 1.50 0.1333
## UniqueAuthors3 0.1816 0.1131 1.61 0.1090
## UniqueAuthors4 0.2750 0.1156 2.38 0.0177 *
## UniqueAuthors5 0.3095 0.1132 2.73 0.0065 **
## Year1997 -0.1668 0.1558 -1.07 0.2850
## Year1998 -0.0859 0.1848 -0.46 0.6424
## Year1999 -0.1793 0.1374 -1.31 0.1925
```

```

## Year2000          -0.2162      0.1480    -1.46    0.1446
## Year2001          -0.2209      0.1625    -1.36    0.1747
## Year2002          -0.2919      0.1323    -2.21    0.0279 *
## Year2003          -0.3504      0.1627    -2.15    0.0317 *
## Year2004          -0.3066      0.1508    -2.03    0.0427 *
## Year2005          -0.3184      0.1432    -2.22    0.0266 *
## Year2006          -0.2623      0.1335    -1.96    0.0501 .
## Year2007          -0.3375      0.1490    -2.27    0.0239 *
## Year2008          -0.1637      0.1331    -1.23    0.2194
## Year2009          -0.3873      0.1421    -2.73    0.0067 **
## Year2010          -0.2566      0.1294    -1.98    0.0480 *
## Year2011          -0.2893      0.1287    -2.25    0.0250 *
## Year2012          -0.3265      0.1310    -2.49    0.0130 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0759, Adjusted R-squared:  0.0341
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 17 is an outlier with |weight| = 0 ( < 0.0002);
## 41 weights are ~= 1. The remaining 467 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0067 0.8610 0.9510 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          1.96e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
## factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.166 1          1.080
## LastAuthorFemale  1.193 1          1.092
## Year              1.386 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
## -1.2941 -0.2541 0.0117 0.2648 1.8904
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45360 0.10919 13.31 <2e-16 ***
## FirstAuthorFemale1 0.04341 0.04246 1.02 0.307
## LastAuthorFemale1 -0.00777 0.07001 -0.11 0.912
## Year1997 -0.15949 0.14886 -1.07 0.285
## Year1998 -0.10888 0.17246 -0.63 0.528
## Year1999 -0.16202 0.13008 -1.25 0.214
## Year2000 -0.18942 0.14271 -1.33 0.185
## Year2001 -0.22859 0.15337 -1.49 0.137
## Year2002 -0.28449 0.12506 -2.27 0.023 *
## Year2003 -0.32247 0.14995 -2.15 0.032 *
## Year2004 -0.28478 0.14624 -1.95 0.052 .
## Year2005 -0.29175 0.13467 -2.17 0.031 *
```

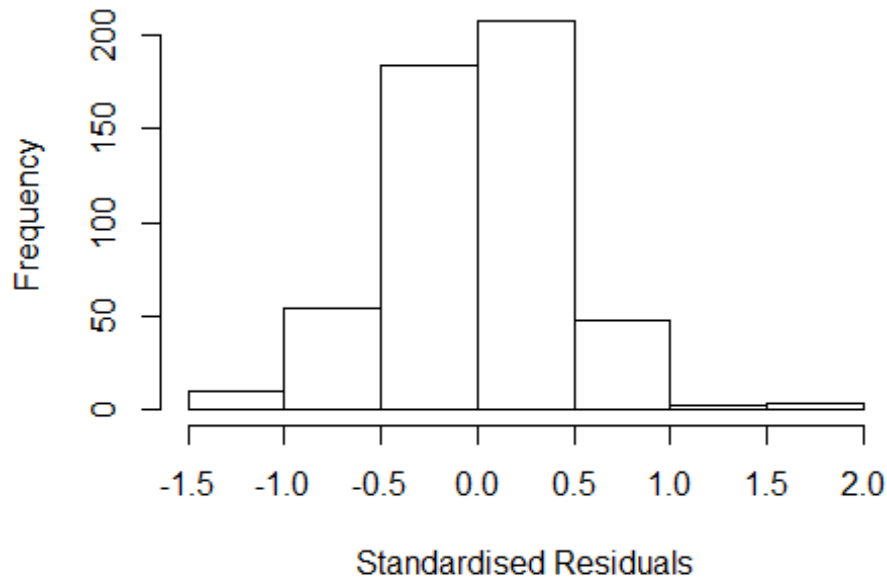


```

## Year2006      -0.24776    0.12337   -2.01    0.045 *
## Year2007      -0.28377    0.14508   -1.96    0.051 .
## Year2008      -0.13252    0.12594   -1.05    0.293
## Year2009      -0.32589    0.13239   -2.46    0.014 *
## Year2010      -0.19649    0.11982   -1.64    0.102
## Year2011      -0.24813    0.12134   -2.04    0.041 *
## Year2012      -0.28451    0.12457   -2.28    0.023 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0375, Adjusted R-squared:  0.00219
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 17 is an outlier with |weight| = 0 ( < 0.0002);
## 43 weights are ~= 1. The remaining 465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0299 0.8520 0.9480 0.8900 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      1.96e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.17 1          1.082
## Year              1.17 16          1.005

```

Residuals from first author



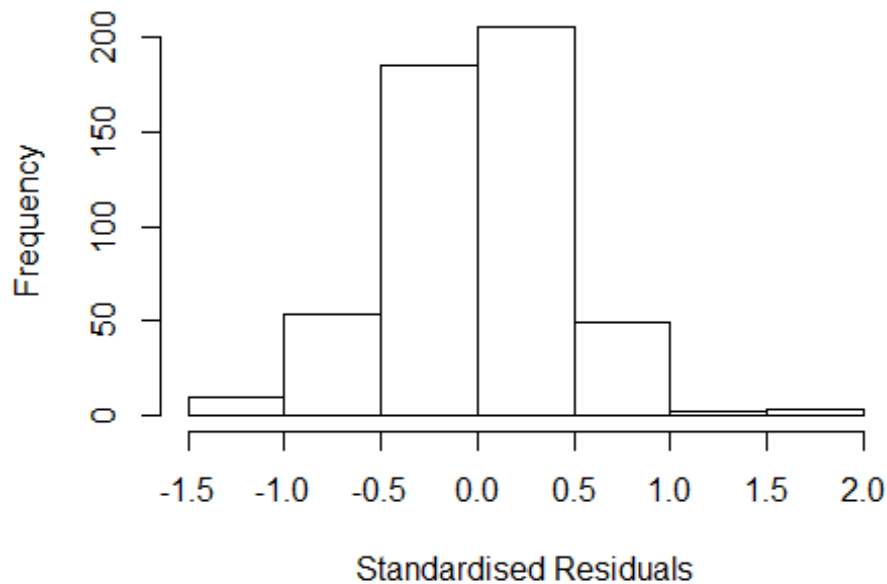
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2944 -0.2531 0.0137 0.2634 1.8904
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4536 0.1091 13.32 <2e-16 ***
## FirstAuthorFemale1 0.0430 0.0427 1.01 0.315
## Year1997 -0.1592 0.1488 -1.07 0.285
## Year1998 -0.1095 0.1722 -0.64 0.525
## Year1999 -0.1628 0.1301 -1.25 0.212
## Year2000 -0.1891 0.1427 -1.32 0.186
## Year2001 -0.2271 0.1525 -1.49 0.137
## Year2002 -0.2860 0.1242 -2.30 0.022 *
## Year2003 -0.3230 0.1500 -2.15 0.032 *
## Year2004 -0.2855 0.1465 -1.95 0.052 .
## Year2005 -0.2932 0.1342 -2.18 0.029 *
## Year2006 -0.2480 0.1232 -2.01 0.045 *
```

```

## Year2007          -0.2840      0.1451   -1.96    0.051 .
## Year2008          -0.1332      0.1255   -1.06    0.289
## Year2009          -0.3263      0.1323   -2.47    0.014 *
## Year2010          -0.1978      0.1189   -1.66    0.097 .
## Year2011          -0.2491      0.1207   -2.06    0.039 *
## Year2012          -0.2855      0.1244   -2.30    0.022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.0375, Adjusted R-squared:  0.00419
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 17 is an outlier with |weight| = 0 ( < 0.0002);
## 43 weights are ~= 1. The remaining 465 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0278 0.8520 0.9480 0.8890 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          1.96e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.193 1          1.092
## Year            1.193 16          1.006

```

Residuals from last author



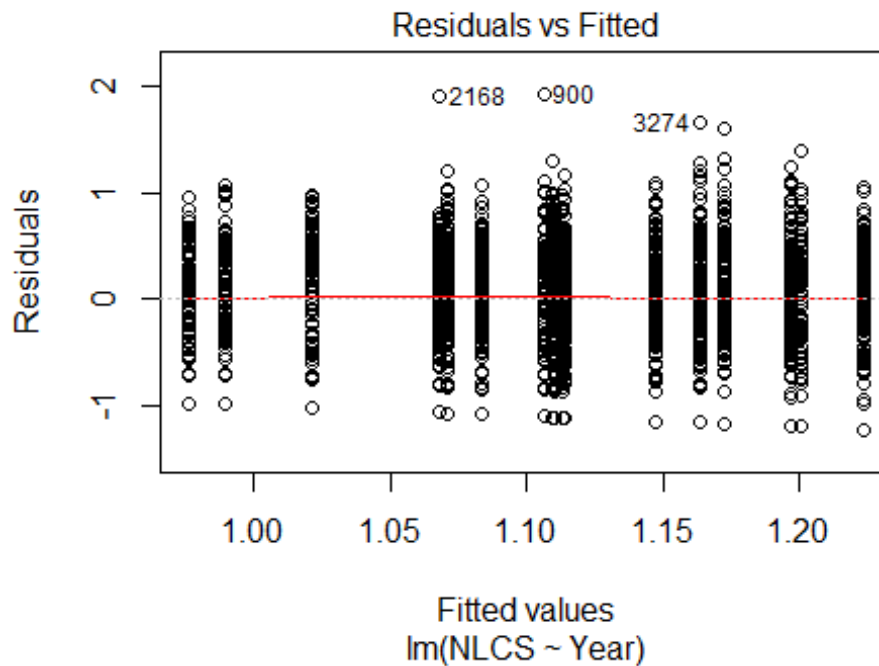
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3059 -0.2550 0.0195 0.2508 1.8820
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.46201 0.10696 13.67 <2e-16 ***
## LastAuthorFemale1 -0.00385 0.07076 -0.05 0.957
## Year1997 -0.15609 0.14762 -1.06 0.291
## Year1998 -0.11384 0.17179 -0.66 0.508
## Year1999 -0.16523 0.12904 -1.28 0.201
## Year2000 -0.18355 0.14207 -1.29 0.197
## Year2001 -0.22853 0.15108 -1.51 0.131
## Year2002 -0.28742 0.12373 -2.32 0.021 *
## Year2003 -0.30947 0.14766 -2.10 0.037 *
## Year2004 -0.28433 0.14493 -1.96 0.050 .
## Year2005 -0.28381 0.13387 -2.12 0.035 *
## Year2006 -0.24319 0.12202 -1.99 0.047 *
```

```

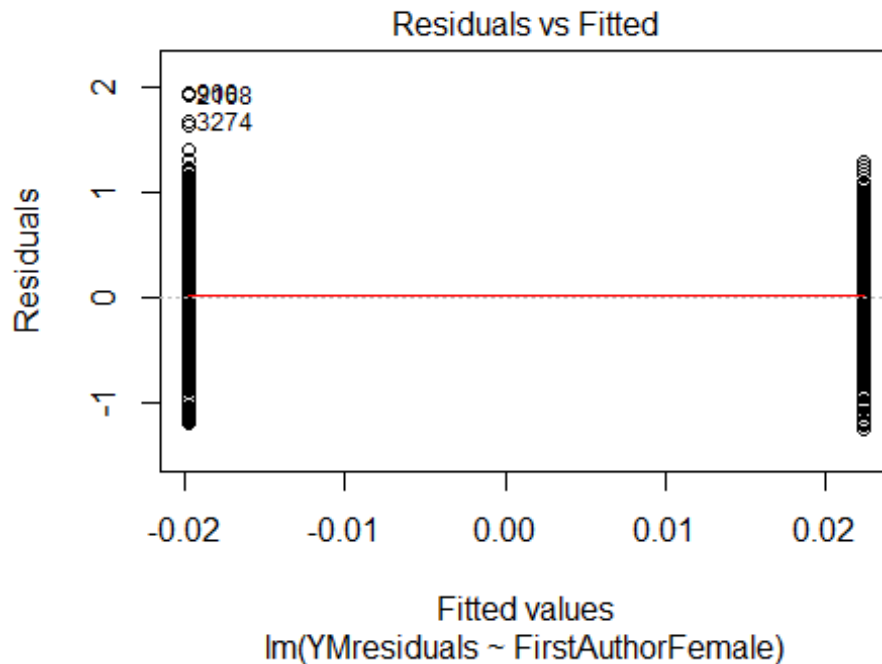
## Year2007          -0.28055      0.14579    -1.92      0.055 .
## Year2008          -0.12224      0.12441    -0.98      0.326
## Year2009          -0.32156      0.13183    -2.44      0.015 *
## Year2010          -0.19249      0.11850    -1.62      0.105
## Year2011          -0.23845      0.11956    -1.99      0.047 *
## Year2012          -0.28100      0.12332    -2.28      0.023 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0359, Adjusted R-squared:  0.00257
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 17 is an outlier with |weight| = 0 ( < 0.0002);
## 36 weights are ~= 1. The remaining 472 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0237 0.8520 0.9530 0.8910 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          1.96e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 509"
## [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
## 117 99 125 116 124 141 168 149 169 167 201 233 212 190 212
## 2011 2012
## 196 246
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 85 104 96 98 100 142 129 155 145 177 205 191 169 176

```

```
## 2011 2012
## 171 217
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 101 76 94 90 89 96 135 116 149 138 162 191 177 161 159
## 2011 2012
## 154 199
## [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
```

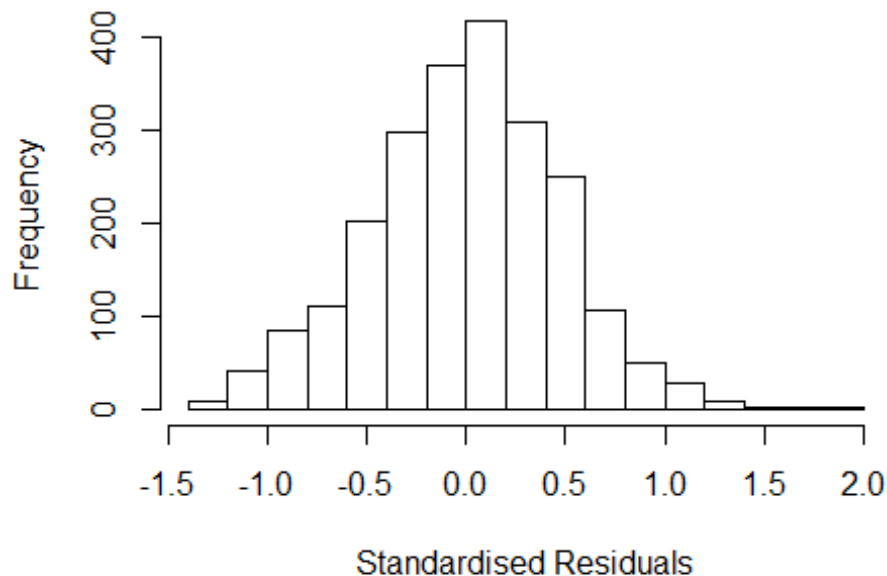


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.86929442453065"
## [1] "Male first author team size 2018 geometric mean: 2.43631494898705"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.89309611584409"
## [1] "Male last author team size 2018 geometric mean: 2.47811184365522"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, 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 1.119 1      1.058
## LastAuthorFemale  1.089 1      1.044
## UniqueAuthors    1.288 4      1.032
## Year              1.300 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
## -1.2991 -0.3013  0.0144  0.3157  1.8869
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92977    0.06647   13.99 < 2e-16 ***
## FirstAuthorFemale1 -0.00726    0.02100   -0.35  0.72945
## LastAuthorFemale1  0.06886    0.02083    3.31  0.00096 ***
## UniqueAuthors2    0.19444    0.02955    6.58  5.8e-11 ***
## UniqueAuthors3    0.20644    0.03247    6.36  2.5e-10 ***
## UniqueAuthors4    0.20054    0.03898    5.15  2.9e-07 ***
## UniqueAuthors5    0.28633    0.03829    7.48  1.1e-13 ***
## Year1997          0.06050    0.08657    0.70  0.48473
## Year1998         -0.05245    0.08424   -0.62  0.53356
## Year1999         -0.07488    0.08183   -0.92  0.36028
```

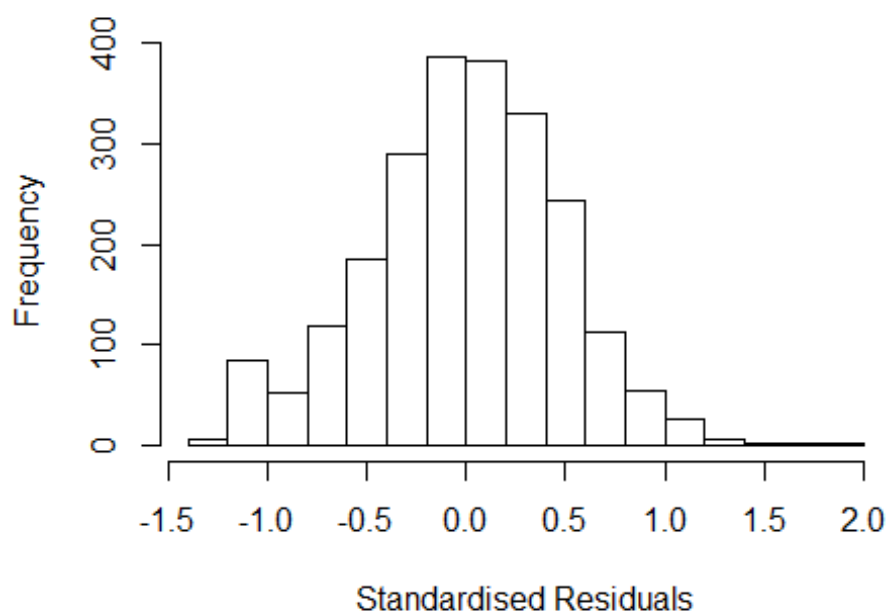


```

## Year2000          0.01027    0.08718    0.12  0.90621
## Year2001          0.03636    0.08613    0.42  0.67297
## Year2002          0.00110    0.08071    0.01  0.98910
## Year2003          0.01247    0.07761    0.16  0.87232
## Year2004          0.09564    0.07716    1.24  0.21532
## Year2005          0.00739    0.07233    0.10  0.91867
## Year2006         -0.01891    0.07256   -0.26  0.79436
## Year2007         -0.05321    0.07080   -0.75  0.45237
## Year2008          0.03847    0.07310    0.53  0.59877
## Year2009          0.02145    0.07799    0.28  0.78330
## Year2010          0.05275    0.07380    0.71  0.47484
## Year2011          0.02070    0.07408    0.28  0.78000
## Year2012          0.01742    0.07259    0.24  0.81038
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0588, Adjusted R-squared:  0.0497
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 2108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0545 0.8720 0.9500 0.9030 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      4.37e-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 1.094 1      1.046
## LastAuthorFemale 1.079 1      1.039
## Year              1.080 16      1.002

```

Residuals from first and last author

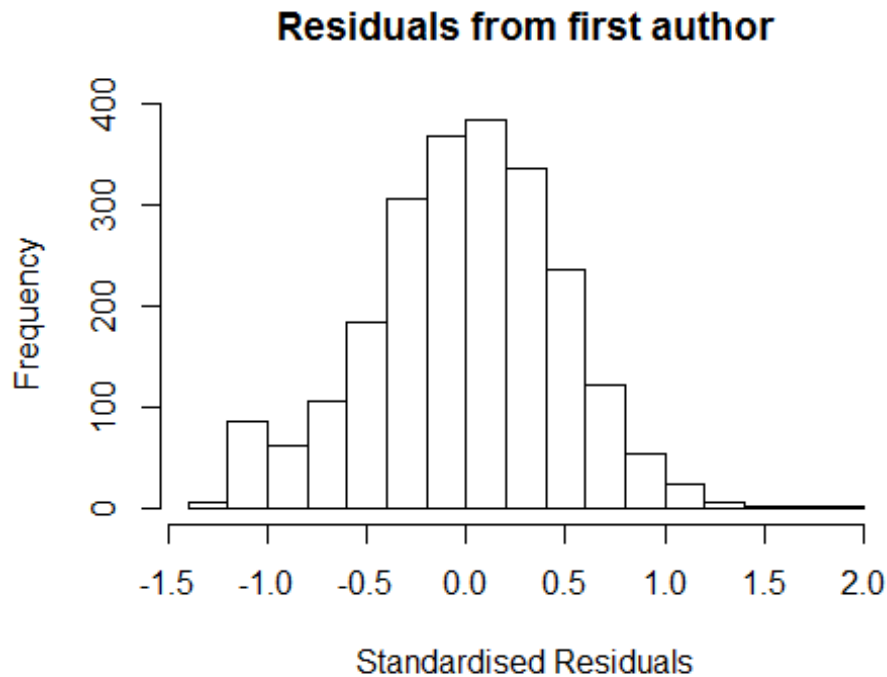


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2879 -0.3060 0.0118 0.3197 1.9560
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0081 0.0678 14.86 < 2e-16 ***
## FirstAuthorFemale1 0.0198 0.0212 0.93 0.35048
## LastAuthorFemale1 0.0814 0.0211 3.85 0.00012 ***
## Year1997 0.0977 0.0908 1.08 0.28198
## Year1998 -0.0598 0.0879 -0.68 0.49634
## Year1999 -0.0417 0.0853 -0.49 0.62500
## Year2000 0.0437 0.0899 0.49 0.62670
## Year2001 0.0629 0.0893 0.70 0.48149
## Year2002 0.0470 0.0845 0.56 0.57815
## Year2003 0.0549 0.0816 0.67 0.50158
## Year2004 0.1786 0.0792 2.25 0.02426 *
## Year2005 0.0767 0.0751 1.02 0.30749
```

```

## Year2006          0.0514      0.0755      0.68  0.49641
## Year2007          0.0197      0.0736      0.27  0.78872
## Year2008          0.1196      0.0755      1.59  0.11310
## Year2009          0.0962      0.0808      1.19  0.23391
## Year2010          0.1273      0.0761      1.67  0.09455 .
## Year2011          0.0953      0.0773      1.23  0.21762
## Year2012          0.0905      0.0759      1.19  0.23312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.0231, Adjusted R-squared:  0.0154
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 2115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0352 0.8680 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      4.37e-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 1.05 1      1.025
## Year              1.05 16      1.002

```

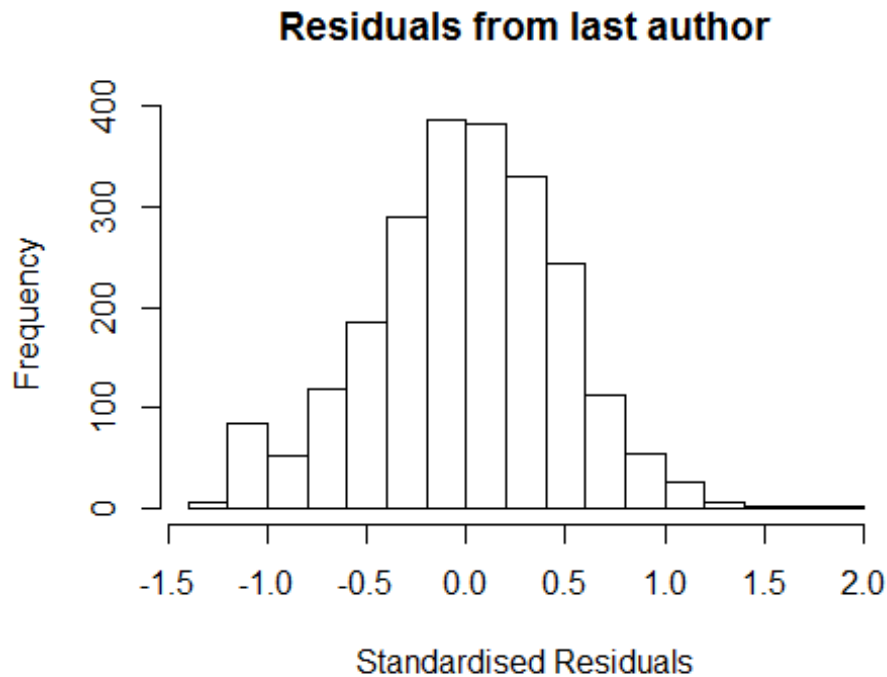


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2538 -0.3112  0.0194  0.3172  1.9378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0339    0.0669   15.46  <2e-16 ***
## FirstAuthorFemale1  0.0433    0.0209    2.07   0.039 *
## Year1997          0.0974    0.0896    1.09   0.277
## Year1998         -0.0709    0.0875   -0.81   0.418
## Year1999         -0.0441    0.0847   -0.52   0.602
## Year2000          0.0370    0.0897    0.41   0.680
## Year2001          0.0552    0.0894    0.62   0.537
## Year2002          0.0369    0.0844    0.44   0.662
## Year2003          0.0541    0.0813    0.67   0.506
## Year2004          0.1766    0.0782    2.26   0.024 *
## Year2005          0.0699    0.0746    0.94   0.349
## Year2006          0.0499    0.0750    0.67   0.505
```

```

## Year2007          0.0198      0.0732      0.27      0.787
## Year2008          0.1193      0.0748      1.59      0.111
## Year2009          0.0966      0.0806      1.20      0.230
## Year2010          0.1276      0.0755      1.69      0.091 .
## Year2011          0.0956      0.0768      1.25      0.213
## Year2012          0.0884      0.0755      1.17      0.242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0169, Adjusted R-squared:  0.00951
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 167 weights are ~= 1. The remaining 2120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0449 0.8690 0.9520 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      4.37e-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 1.038 1      1.019
## Year      1.038 16      1.001

```



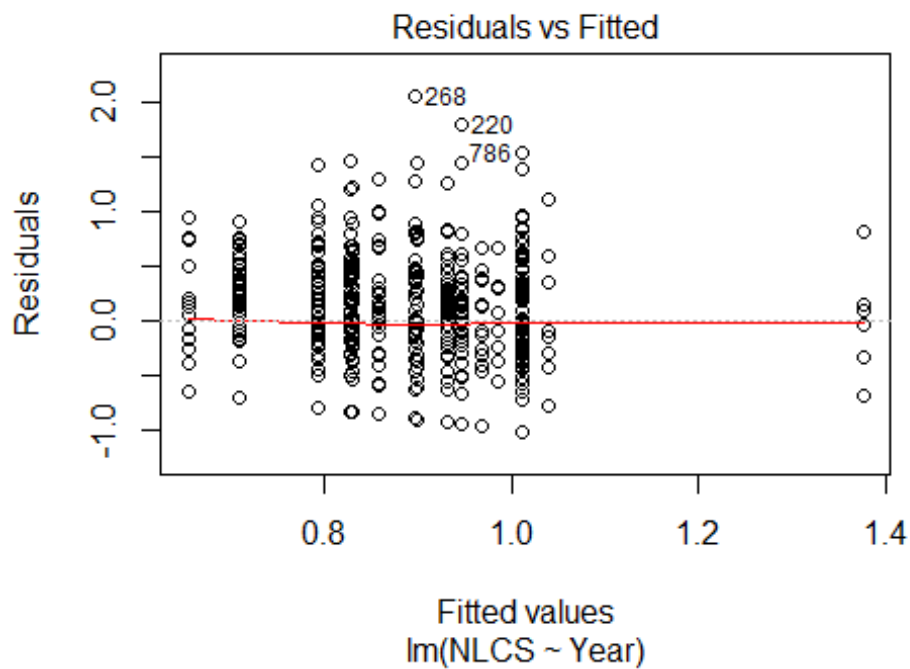
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28115 -0.30695  0.00872  0.32079  1.95026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0131     0.0675   15.00 < 2e-16 ***
## LastAuthorFemale1  0.0871     0.0208    4.19 2.9e-05 ***
## Year1997          0.0981     0.0908    1.08  0.280
## Year1998         -0.0605     0.0877   -0.69  0.490
## Year1999         -0.0417     0.0851   -0.49  0.624
## Year2000          0.0444     0.0897    0.49  0.621
## Year2001          0.0636     0.0890    0.72  0.475
## Year2002          0.0485     0.0842    0.58  0.565
## Year2003          0.0570     0.0812    0.70  0.483
## Year2004          0.1809     0.0788    2.29  0.022 *
## Year2005          0.0787     0.0747    1.05  0.292
## Year2006          0.0539     0.0750    0.72  0.472
```

```

## Year2007          0.0212      0.0733      0.29      0.772
## Year2008          0.1220      0.0751      1.63      0.104
## Year2009          0.1003      0.0802      1.25      0.211
## Year2010          0.1305      0.0757      1.72      0.085 .
## Year2011          0.0989      0.0768      1.29      0.198
## Year2012          0.0942      0.0752      1.25      0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.0154
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 2119 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.037  0.869  0.951  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      4.37e-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: 2287"
## [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
##   22   26   20   27   28   32   40   46   43   34   36   49   56   75   65
## 2011 2012
##   56   60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21   21    9   14    9   25    6   41   39   33   33   45   52   67   59
## 2011 2012

```

```
## 48 55
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 20 7 14 8 24 5 38 37 33 30 45 49 66 57
## 2011 2012
## 45 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 = 0.8
```

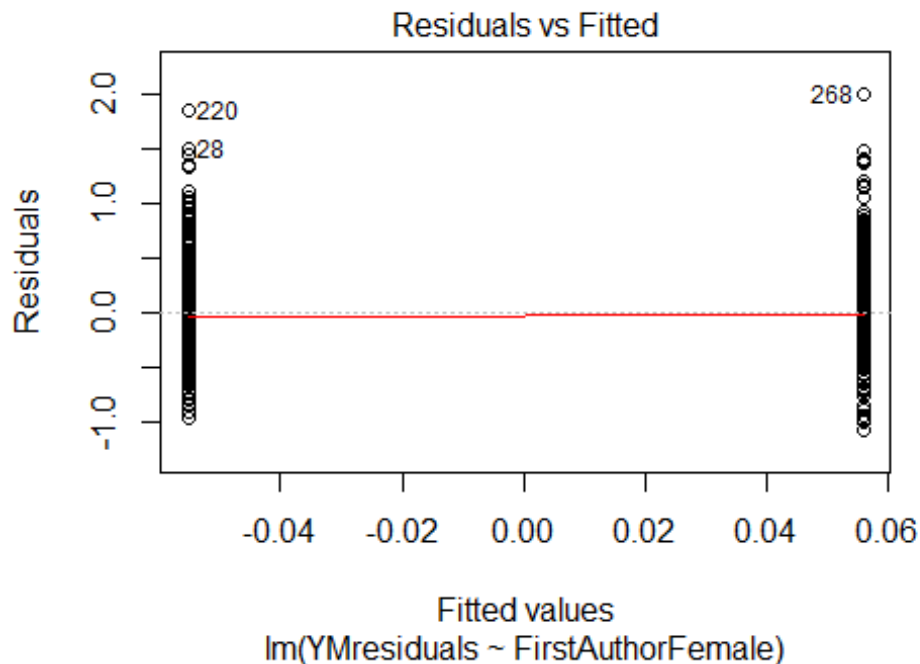


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.053, df = 1, p-value = 0.8
## [1] "Female first author team size 2018 geometric mean: 2.7583164166863"
## [1] "Male first author team size 2018 geometric mean: 2.17223571494358"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



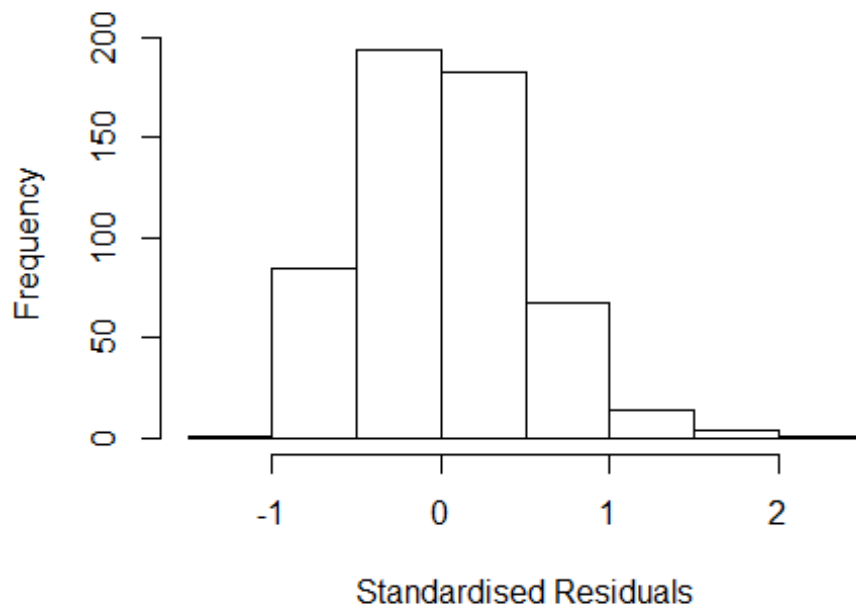
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 230, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.55131179282065"
## [1] "Male last author team size 2018 geometric mean: 2.45885782921044"

## 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 = 210, 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 1.609  1      1.268
## LastAuthorFemale  1.570  1      1.253
## UniqueAuthors    1.726  4      1.071
## Year              1.863 16      1.020
```

Residuals from first and last author and team size



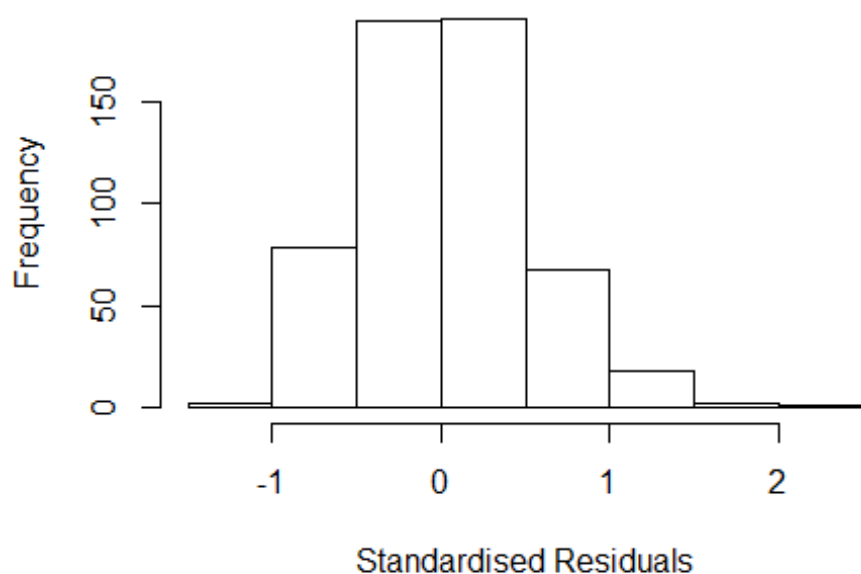
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0706 -0.3183 -0.0072 0.3179 2.2081
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5425 0.1207 4.49 8.6e-06 ***
## FirstAuthorFemale1 0.1270 0.0511 2.48 0.01330 *
## LastAuthorFemale1 -0.0212 0.0512 -0.41 0.67964
## UniqueAuthors2 0.2079 0.0514 4.04 6.1e-05 ***
## UniqueAuthors3 0.1300 0.0692 1.88 0.06081 .
## UniqueAuthors4 0.3092 0.0891 3.47 0.00056 ***
## UniqueAuthors5 0.2192 0.1118 1.96 0.05055 .
## Year1997 0.1426 0.1702 0.84 0.40249
## Year1998 0.3996 0.1548 2.58 0.01011 *
## Year1999 0.2874 0.1736 1.66 0.09843 .
```

```

## Year2000          0.2682      0.2656      1.01  0.31296
## Year2001          0.0673      0.1557      0.43  0.66595
## Year2002          0.7659      0.2323      3.30  0.00104 **
## Year2003          0.1608      0.1418      1.13  0.25746
## Year2004          0.0906      0.1466      0.62  0.53676
## Year2005          0.2362      0.1482      1.59  0.11156
## Year2006          0.1471      0.1484      0.99  0.32201
## Year2007          0.1111      0.1352      0.82  0.41146
## Year2008          0.1932      0.1321      1.46  0.14406
## Year2009          0.1183      0.1278      0.93  0.35486
## Year2010          0.0155      0.1345      0.12  0.90822
## Year2011          0.0789      0.1426      0.55  0.57997
## Year2012          0.2397      0.1339      1.79  0.07390 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.474
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0739
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 142 is an outlier with |weight| <= 0.00014 ( < 0.00018);
## 50 weights are ~= 1. The remaining 499 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.861  0.950  0.902  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-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.433 1 1.197
## LastAuthorFemale 1.439 1 1.199
## Year 1.159 16 1.005

```

Residuals from first and last author



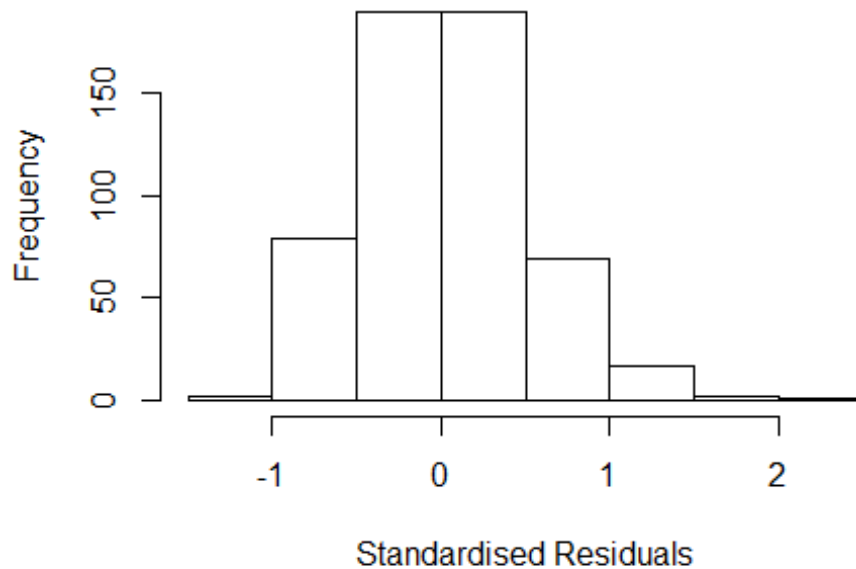
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.05127 -0.30601 0.00669 0.31800 2.09365
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5731 0.1155 4.96 9.5e-07 ***
## FirstAuthorFemale1 0.1522 0.0503 3.02 0.00261 **
## LastAuthorFemale1 -0.0296 0.0510 -0.58 0.56234
## Year1997 0.1883 0.1760 1.07 0.28523
## Year1998 0.4684 0.1615 2.90 0.00387 **
## Year1999 0.3713 0.1633 2.27 0.02340 *
## Year2000 0.4019 0.2540 1.58 0.11413
## Year2001 0.1423 0.1574 0.90 0.36623
## Year2002 0.8217 0.2321 3.54 0.00043 ***
## Year2003 0.2310 0.1379 1.67 0.09459 .
## Year2004 0.1576 0.1533 1.03 0.30449
## Year2005 0.3555 0.1420 2.50 0.01260 *
```

```

## Year2006          0.2237      0.1500      1.49  0.13652
## Year2007          0.1878      0.1298      1.45  0.14840
## Year2008          0.2673      0.1298      2.06  0.04002 *
## Year2009          0.1639      0.1251      1.31  0.19096
## Year2010          0.0629      0.1302      0.48  0.62941
## Year2011          0.1632      0.1377      1.19  0.23632
## Year2012          0.3542      0.1282      2.76  0.00593 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.0726, Adjusted R-squared:  0.0412
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 500 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0192 0.8510 0.9510 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      1.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



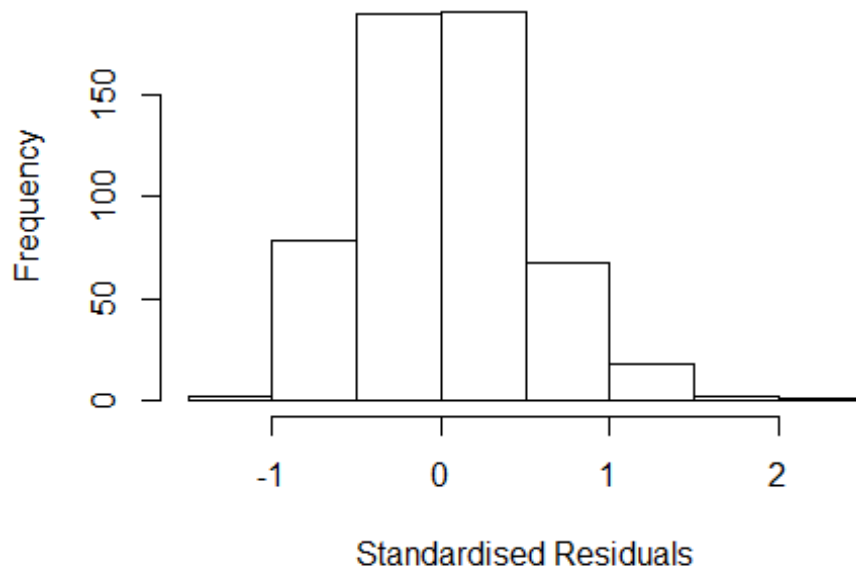
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06248 -0.30009  0.00958  0.32017  2.08198
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5686    0.1152   4.93 1.1e-06 ***
## FirstAuthorFemale1 0.1372    0.0437   3.14  0.0018 **
## Year1997        0.1849    0.1772   1.04  0.2973
## Year1998        0.4692    0.1595   2.94  0.0034 **
## Year1999        0.3710    0.1638   2.27  0.0239 *
## Year2000        0.3897    0.2552   1.53  0.1273
## Year2001        0.1405    0.1573   0.89  0.3722
## Year2002        0.8207    0.2302   3.56  0.0004 ***
## Year2003        0.2301    0.1380   1.67  0.0959 .
## Year2004        0.1592    0.1530   1.04  0.2987
## Year2005        0.3566    0.1416   2.52  0.0121 *
## Year2006        0.2226    0.1501   1.48  0.1387
```

```

## Year2007          0.1890      0.1297      1.46      0.1458
## Year2008          0.2648      0.1300      2.04      0.0421 *
## Year2009          0.1616      0.1252      1.29      0.1975
## Year2010          0.0594      0.1303      0.46      0.6488
## Year2011          0.1646      0.1377      1.20      0.2324
## Year2012          0.3536      0.1282      2.76      0.0060 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.0721, Adjusted R-squared:  0.0425
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 492 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0212 0.8510 0.9480 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      1.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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      1.040
## 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
## -1.02921 -0.32638 0.00368 0.34464 2.11951
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6211 0.1088 5.71 1.9e-08 ***
## LastAuthorFemale1 0.0500 0.0447 1.12 0.2638
## Year1997 0.1733 0.1765 0.98 0.3266
## Year1998 0.4403 0.1638 2.69 0.0074 **
## Year1999 0.3460 0.1635 2.12 0.0348 *
## Year2000 0.3493 0.2530 1.38 0.1679
## Year2001 0.1246 0.1583 0.79 0.4315
## Year2002 0.8010 0.2399 3.34 0.0009 ***
## Year2003 0.2272 0.1327 1.71 0.0874 .
## Year2004 0.1564 0.1489 1.05 0.2942
## Year2005 0.3501 0.1374 2.55 0.0111 *
## Year2006 0.2213 0.1491 1.48 0.1382
```

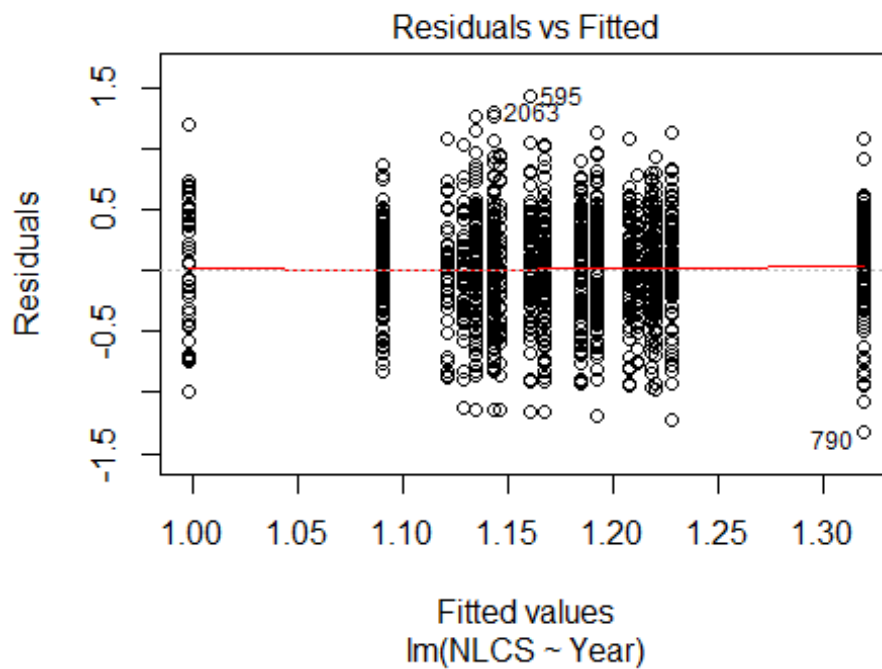


```

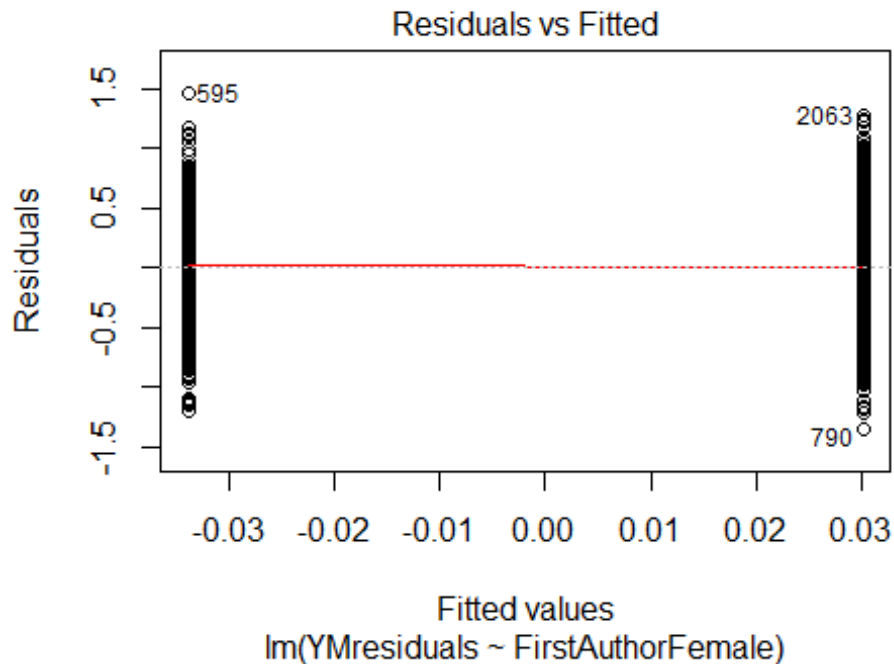
## Year2007          0.1812      0.1275      1.42      0.1557
## Year2008          0.2719      0.1251      2.17      0.0302 *
## Year2009          0.1579      0.1213      1.30      0.1935
## Year2010          0.0496      0.1267      0.39      0.6958
## Year2011          0.1585      0.1350      1.17      0.2409
## Year2012          0.3581      0.1241      2.88      0.0041 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.492
## Multiple R-squared:  0.0555, Adjusted R-squared:  0.0253
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0244 0.8620 0.9480 0.9020 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      1.82e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 550"
## [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
##   57   76   68   60   89  110  109   78   95   81  113  111  127  130  135
## 2011 2012
##  152  156
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   50   54   49   51   81   82   64   82   75  100   98  110  121  109
## 2011 2012

```

```
## 130 121
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 45 48 48 49 79 74 56 69 72 93 91 103 113 96
## 2011 2012
## 116 113
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 6e-04
```

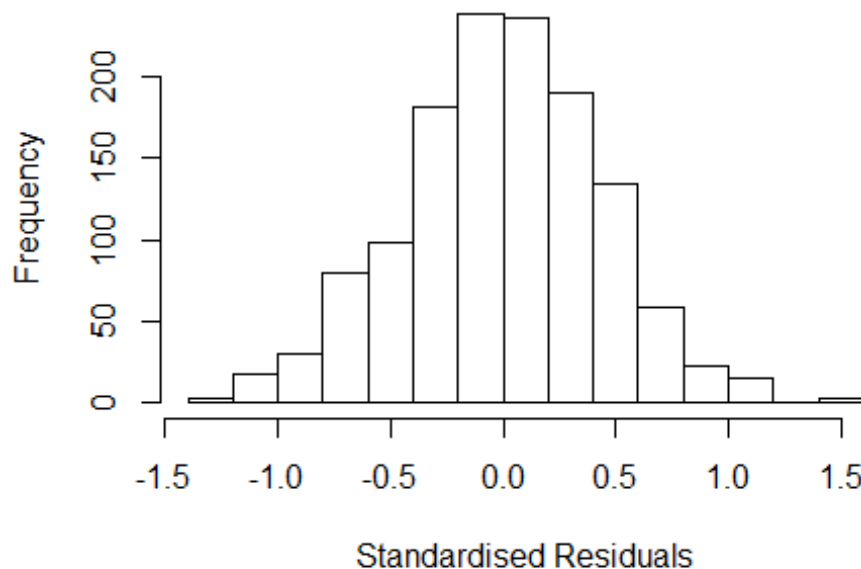


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.33, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 3.00127149810612"
## [1] "Male first author team size 2018 geometric mean: 2.52342733742074"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5800, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.71835781952687"
## [1] "Male last author team size 2018 geometric mean: 2.87284751664232"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4900, 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 1.142 1      1.069
## LastAuthorFemale  1.155 1      1.075
## UniqueAuthors     1.408 4      1.044
## Year               1.444 16     1.012
```

Residuals from first and last author and team size



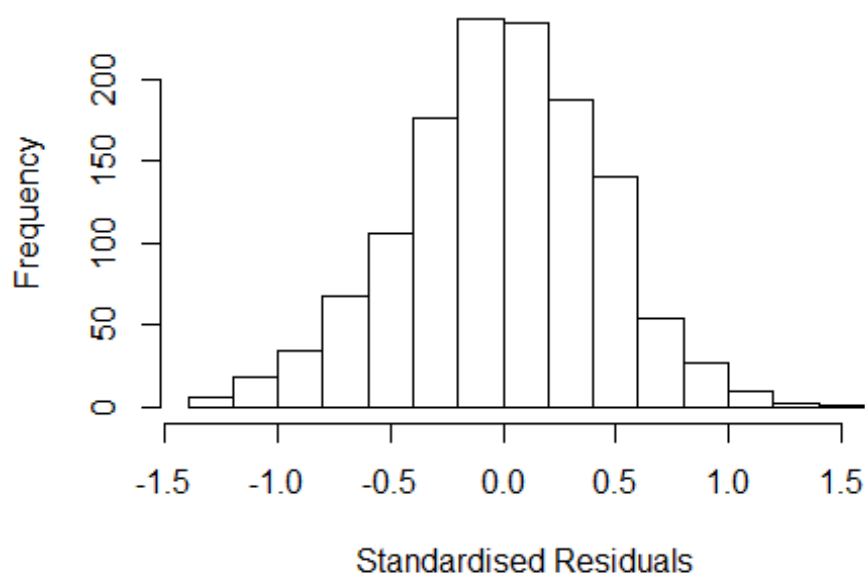
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38251 -0.28638 0.00331 0.29902 1.59469
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9088 0.1078 8.43 < 2e-16 ***
## FirstAuthorFemale1 0.0435 0.0261 1.67 0.0952 .
## LastAuthorFemale1 0.0192 0.0257 0.75 0.4551
## UniqueAuthors2 0.1376 0.0387 3.55 0.0004 ***
## UniqueAuthors3 0.1899 0.0396 4.79 1.8e-06 ***
## UniqueAuthors4 0.2251 0.0497 4.53 6.5e-06 ***
## UniqueAuthors5 0.1742 0.0558 3.12 0.0018 **
## Year1997 0.0869 0.1234 0.70 0.4815
## Year1998 0.0661 0.1248 0.53 0.5966
## Year1999 0.1775 0.1193 1.49 0.1372
```

```

## Year2000          0.1193      0.1354      0.88      0.3782
## Year2001          0.2142      0.1208      1.77      0.0765 .
## Year2002          0.0835      0.1202      0.69      0.4876
## Year2003          0.1448      0.1234      1.17      0.2407
## Year2004          0.2733      0.1200      2.28      0.0229 *
## Year2005          0.1490      0.1169      1.27      0.2028
## Year2006          0.1392      0.1154      1.21      0.2278
## Year2007          0.0242      0.1129      0.21      0.8306
## Year2008          0.0902      0.1157      0.78      0.4360
## Year2009          0.0580      0.1171      0.50      0.6202
## Year2010          0.0308      0.1202      0.26      0.7976
## Year2011          0.1253      0.1168      1.07      0.2836
## Year2012          0.0215      0.1168      0.18      0.8542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0584, Adjusted R-squared:  0.0422
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.136  0.876  0.950  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      7.66e-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 1.110 1      1.054
## LastAuthorFemale  1.146 1      1.070
## Year              1.085 16      1.003

```

Residuals from first and last author



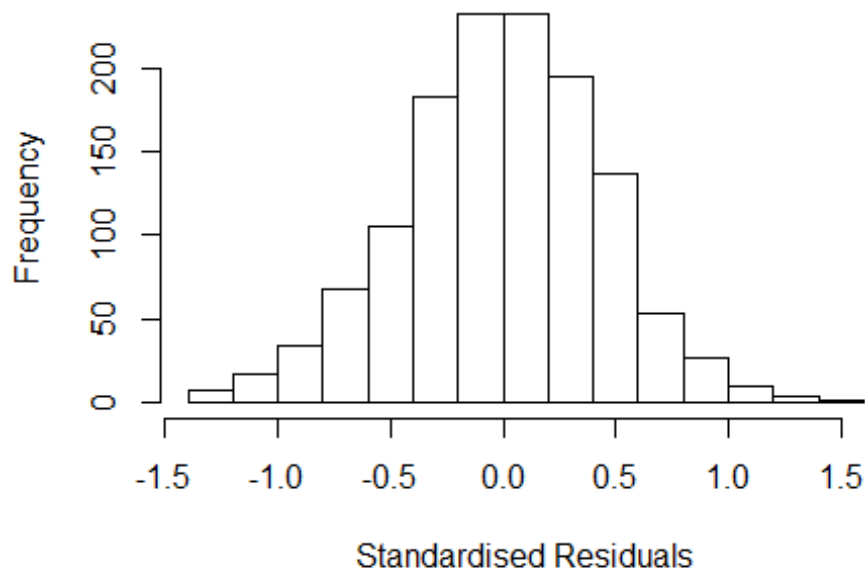
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39162 -0.27723 0.00239 0.29939 1.46205
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9651 0.1127 8.57 <2e-16 ***
## FirstAuthorFemale1 0.0625 0.0260 2.40 0.0165 *
## LastAuthorFemale1 0.0213 0.0260 0.82 0.4136
## Year1997 0.1103 0.1284 0.86 0.3907
## Year1998 0.1232 0.1278 0.96 0.3352
## Year1999 0.2105 0.1224 1.72 0.0856 .
## Year2000 0.1481 0.1377 1.08 0.2825
## Year2001 0.2609 0.1251 2.09 0.0371 *
## Year2002 0.1598 0.1234 1.30 0.1955
## Year2003 0.2015 0.1257 1.60 0.1092
## Year2004 0.3427 0.1239 2.77 0.0057 **
## Year2005 0.2038 0.1213 1.68 0.0932 .
```

```

## Year2006          0.2229      0.1189      1.87      0.0611 .
## Year2007          0.0762      0.1178      0.65      0.5181
## Year2008          0.1651      0.1198      1.38      0.1685
## Year2009          0.1385      0.1206      1.15      0.2512
## Year2010          0.1013      0.1242      0.82      0.4148
## Year2011          0.2047      0.1207      1.70      0.0900 .
## Year2012          0.1070      0.1209      0.89      0.3760
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0312, Adjusted R-squared:  0.0176
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1195 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.232  0.876  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      7.66e-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 1.034 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
## -1.38210 -0.28283 0.00781 0.30064 1.45501
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9698 0.1122 8.65 <2e-16 ***
## FirstAuthorFemale1 0.0697 0.0252 2.77 0.0057 **
## Year1997 0.1113 0.1281 0.87 0.3851
## Year1998 0.1250 0.1273 0.98 0.3266
## Year1999 0.2108 0.1219 1.73 0.0841 .
## Year2000 0.1504 0.1373 1.10 0.2735
## Year2001 0.2611 0.1245 2.10 0.0362 *
## Year2002 0.1622 0.1228 1.32 0.1870
## Year2003 0.2040 0.1254 1.63 0.1041
## Year2004 0.3426 0.1234 2.78 0.0056 **
## Year2005 0.2057 0.1208 1.70 0.0888 .
## Year2006 0.2251 0.1182 1.91 0.0570 .
```

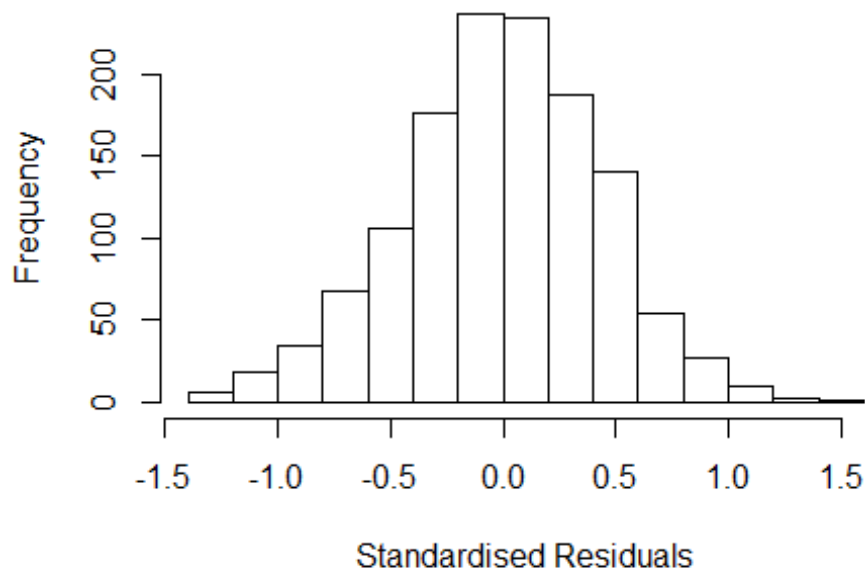


```

## Year2007          0.0758      0.1174      0.65      0.5184
## Year2008          0.1669      0.1194      1.40      0.1623
## Year2009          0.1405      0.1201      1.17      0.2424
## Year2010          0.1019      0.1238      0.82      0.4107
## Year2011          0.2059      0.1203      1.71      0.0871 .
## Year2012          0.1082      0.1204      0.90      0.3691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0307, Adjusted R-squared:  0.0179
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1198 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.236  0.875   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      7.66e-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 1.064 1      1.031
## Year      1.064 16      1.002

```

Residuals from last author



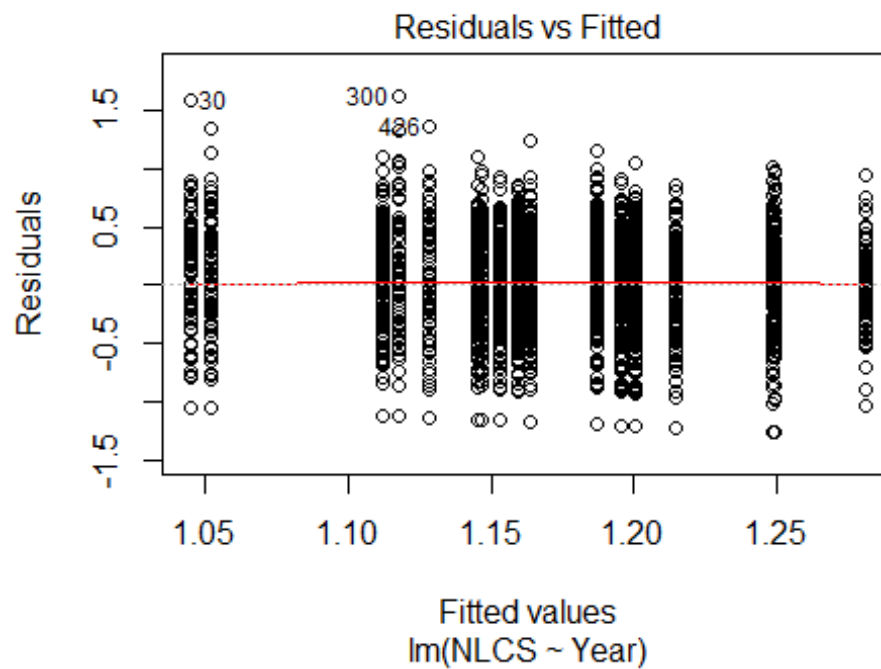
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37583 -0.28633  0.00088  0.30388  1.43667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9932     0.1113   8.92  <2e-16 ***
## LastAuthorFemale1  0.0425     0.0251   1.69   0.091 .
## Year1997          0.1014     0.1274   0.80   0.426
## Year1998          0.1187     0.1272   0.93   0.351
## Year1999          0.1991     0.1216   1.64   0.102
## Year2000          0.1376     0.1376   1.00   0.317
## Year2001          0.2557     0.1252   2.04   0.041 *
## Year2002          0.1571     0.1228   1.28   0.201
## Year2003          0.1972     0.1251   1.58   0.115
## Year2004          0.3401     0.1235   2.75   0.006 **
## Year2005          0.1999     0.1205   1.66   0.097 .
## Year2006          0.2189     0.1183   1.85   0.064 .
```

```

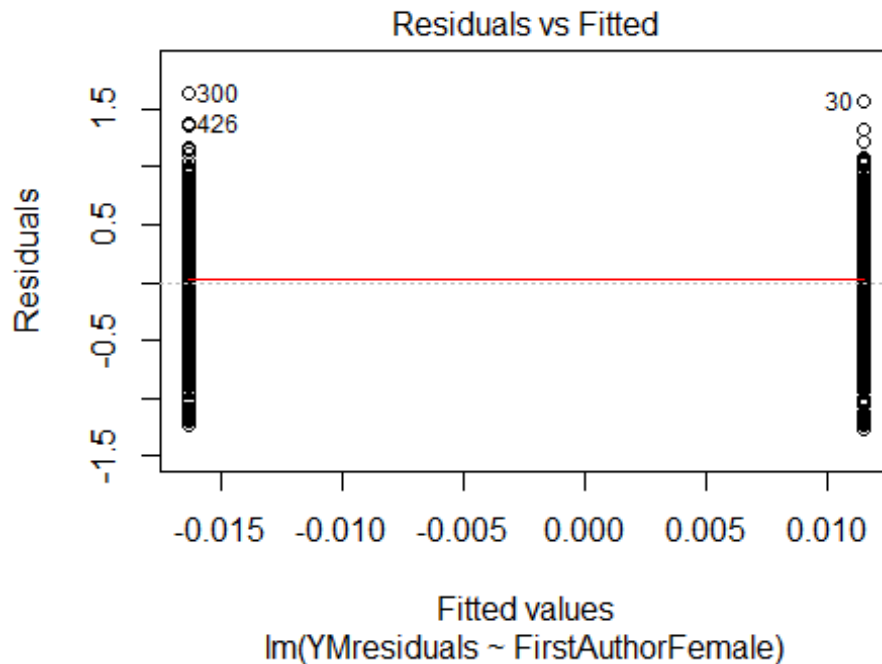
## Year2007          0.0741      0.1173      0.63      0.528
## Year2008          0.1633      0.1193      1.37      0.171
## Year2009          0.1336      0.1198      1.11      0.265
## Year2010          0.0961      0.1231      0.78      0.435
## Year2011          0.1996      0.1200      1.66      0.097 .
## Year2012          0.1053      0.1203      0.88      0.382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0269, Adjusted R-squared:  0.0141
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1192 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.876  0.950  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      7.66e-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: 1305"
## [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
##  100  111  112  106  115  150  131  126  128  157  216  210  197  254  249
## 2011 2012
##  273  293
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   79   81   86   83   73   87  110  105  114  144  182  190  181  218  215
## 2011 2012

```

```
## 232 263
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 75 76 77 75 70 82 99 94 100 132 171 172 165 207 192
## 2011 2012
## 219 241
## [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
```

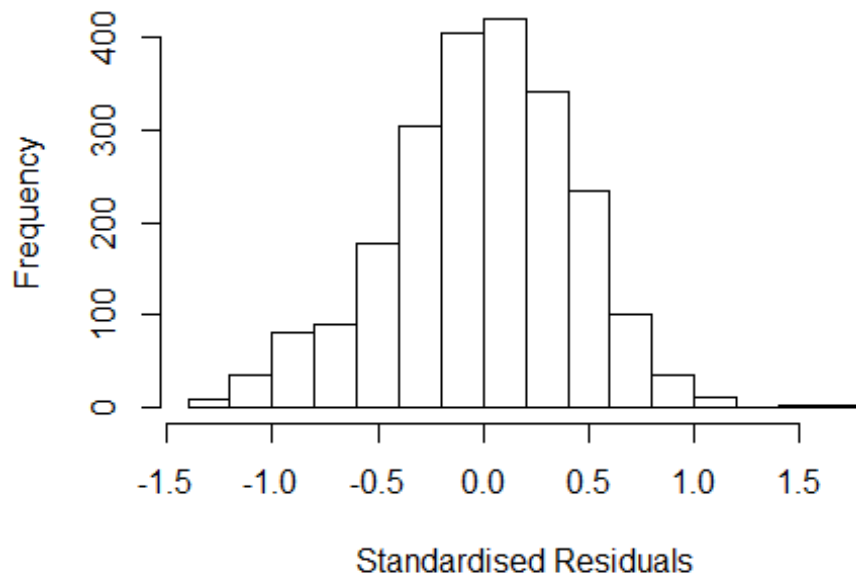


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.61843479178468"
## [1] "Male first author team size 2018 geometric mean: 3.67050371620269"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.54818876665269"
## [1] "Male last author team size 2018 geometric mean: 3.74192805577298"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, 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 1.107 1      1.052
## LastAuthorFemale  1.047 1      1.023
## UniqueAuthors    1.222 4      1.025
## Year             1.290 16      1.008
```

Residuals from first and last author and team size



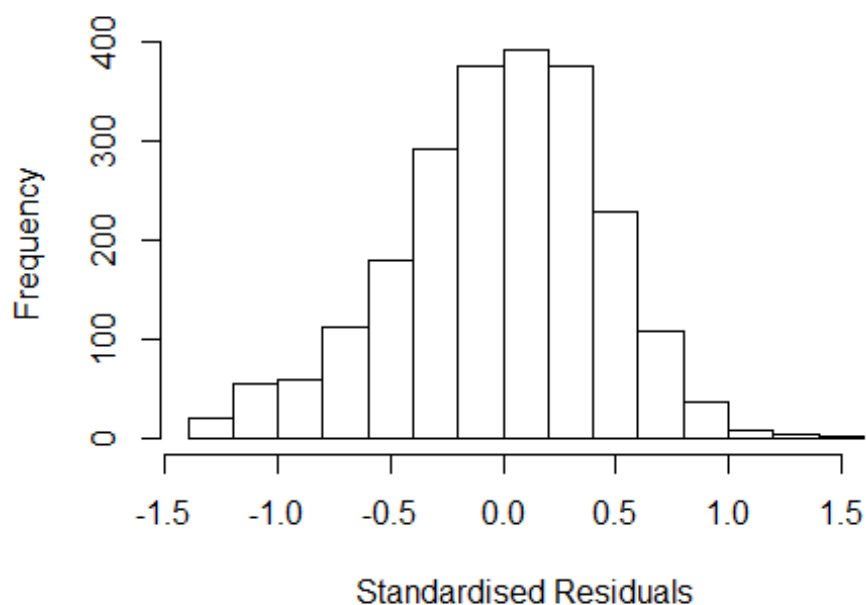
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3577 -0.2827  0.0127  0.2861  1.6250
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.892061   0.072398  12.32  < 2e-16 ***
## FirstAuthorFemale1 0.013315   0.019849   0.67    0.50
## LastAuthorFemale1 0.000638   0.018861   0.03    0.97
## UniqueAuthors2    0.280614   0.034799   8.06  1.2e-15 ***
## UniqueAuthors3    0.269774   0.036244   7.44  1.4e-13 ***
## UniqueAuthors4    0.368851   0.039321   9.38  < 2e-16 ***
## UniqueAuthors5    0.446095   0.037049  12.04  < 2e-16 ***
## Year1997          -0.095370   0.093232  -1.02    0.31
## Year1998          -0.018270   0.103923  -0.18    0.86
## Year1999          -0.033055   0.086976  -0.38    0.70
```

```

## Year2000      0.117477    0.083367    1.41    0.16
## Year2001      0.104798    0.085677    1.22    0.22
## Year2002      0.006243    0.083713    0.07    0.94
## Year2003      0.072050    0.085628    0.84    0.40
## Year2004      0.077909    0.083885    0.93    0.35
## Year2005      0.011856    0.078231    0.15    0.88
## Year2006     -0.020193    0.075847   -0.27    0.79
## Year2007     -0.004982    0.077992   -0.06    0.95
## Year2008     -0.033681    0.076049   -0.44    0.66
## Year2009     -0.053711    0.075855   -0.71    0.48
## Year2010      0.022940    0.075426    0.30    0.76
## Year2011      0.002833    0.074868    0.04    0.97
## Year2012     -0.000859    0.075633   -0.01    0.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0974, Adjusted R-squared:  0.0884
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~= 1. The remaining 2085 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.107  0.876  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      4.45e-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 1.083 1      1.041
## LastAuthorFemale  1.035 1      1.017
## Year              1.092 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
## -1.2880 -0.3024  0.0184  0.3013  1.5347
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06999    0.07474   14.32  <2e-16 ***
## FirstAuthorFemale1  0.02436    0.02042    1.19   0.233
## LastAuthorFemale1 -0.00526    0.01958   -0.27   0.788
## Year1997        -0.06408    0.09986   -0.64   0.521
## Year1998         0.02285    0.10105    0.23   0.821
## Year1999         0.03564    0.09287    0.38   0.701
## Year2000         0.19426    0.08586    2.26   0.024 *
## Year2001         0.19889    0.08989    2.21   0.027 *
## Year2002         0.10235    0.08803    1.16   0.245
## Year2003         0.16908    0.08881    1.90   0.057 .
## Year2004         0.16162    0.08711    1.86   0.064 .
## Year2005         0.10956    0.08319    1.32   0.188
```

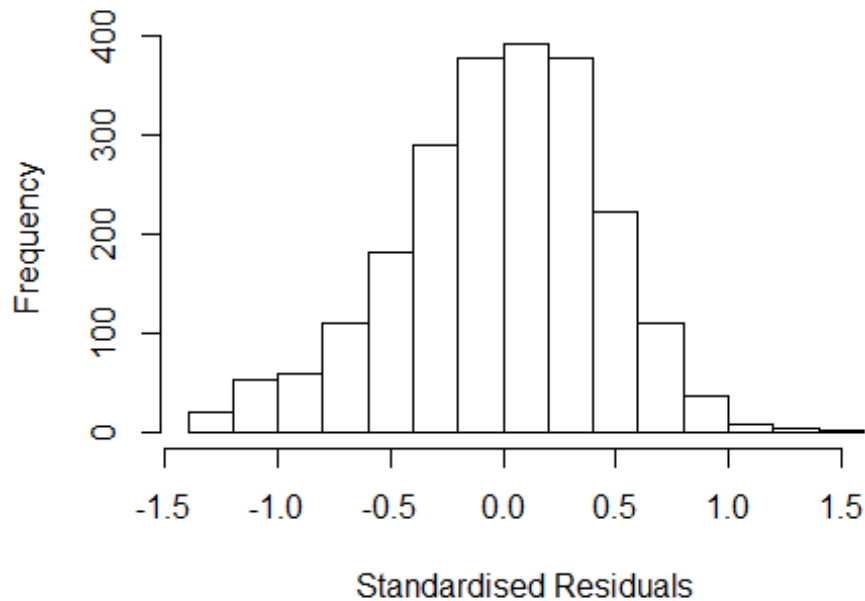


```

## Year2006          0.08250    0.08044    1.03    0.305
## Year2007          0.08433    0.08207    1.03    0.304
## Year2008          0.07842    0.08062    0.97    0.331
## Year2009          0.05265    0.08081    0.65    0.515
## Year2010          0.13798    0.07988    1.73    0.084 .
## Year2011          0.13218    0.07999    1.65    0.099 .
## Year2012          0.13103    0.07965    1.65    0.100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00879
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 2064 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.191  0.873   0.949   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      4.45e-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 1.066 1         1.032
## Year              1.066 16         1.002

```

Residuals from first author



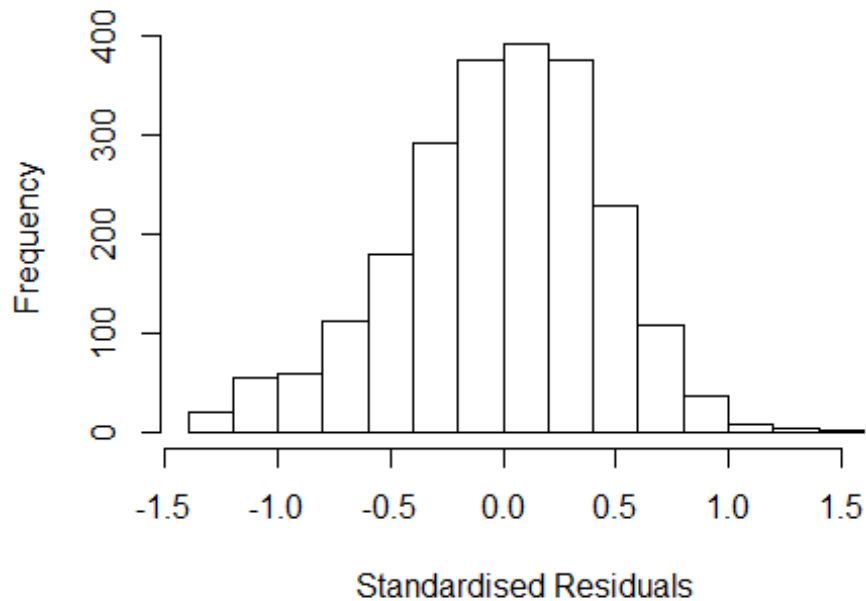
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2900 -0.3036 0.0192 0.3016 1.5372
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0682 0.0747 14.29 <2e-16 ***
## FirstAuthorFemale1 0.0236 0.0203 1.16 0.244
## Year1997 -0.0643 0.0997 -0.65 0.519
## Year1998 0.0227 0.1010 0.22 0.822
## Year1999 0.0358 0.0929 0.39 0.700
## Year2000 0.1938 0.0857 2.26 0.024 *
## Year2001 0.1983 0.0897 2.21 0.027 *
## Year2002 0.1019 0.0879 1.16 0.247
## Year2003 0.1683 0.0885 1.90 0.057 .
## Year2004 0.1621 0.0871 1.86 0.063 .
## Year2005 0.1095 0.0831 1.32 0.188
## Year2006 0.0823 0.0803 1.02 0.306
```

```

## Year2007          0.0840      0.0819      1.03      0.305
## Year2008          0.0782      0.0805      0.97      0.332
## Year2009          0.0524      0.0807      0.65      0.516
## Year2010          0.1377      0.0797      1.73      0.084 .
## Year2011          0.1319      0.0799      1.65      0.099 .
## Year2012          0.1307      0.0795      1.64      0.100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00919
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 181 weights are ~= 1. The remaining 2066 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.874  0.949  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      4.45e-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 1.017 1      1.009
## Year              1.017 16      1.001

```

Residuals from last author



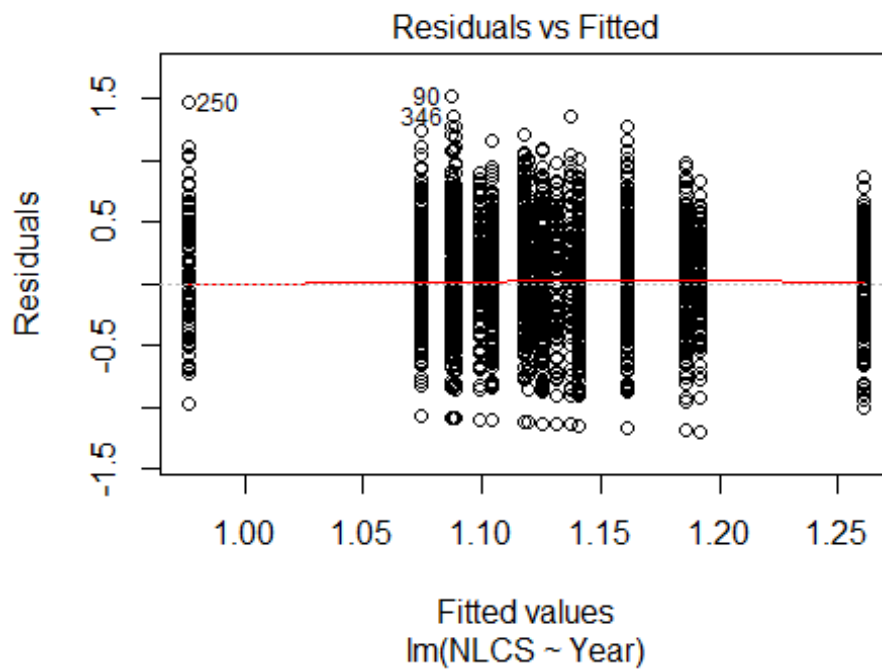
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2795 -0.3056  0.0207  0.2997  1.5505
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07845    0.07378   14.62  <2e-16 ***
## LastAuthorFemale1 -0.00188    0.01943   -0.10    0.923
## Year1997       -0.06409    0.09990   -0.64    0.521
## Year1998        0.02388    0.10090    0.24    0.813
## Year1999        0.03999    0.09255    0.43    0.666
## Year2000        0.19767    0.08565    2.31    0.021 *
## Year2001        0.20100    0.08960    2.24    0.025 *
## Year2002        0.10627    0.08799    1.21    0.227
## Year2003        0.17242    0.08861    1.95    0.052 .
## Year2004        0.16778    0.08687    1.93    0.054 .
## Year2005        0.11636    0.08276    1.41    0.160
## Year2006        0.08790    0.08023    1.10    0.273
```

```

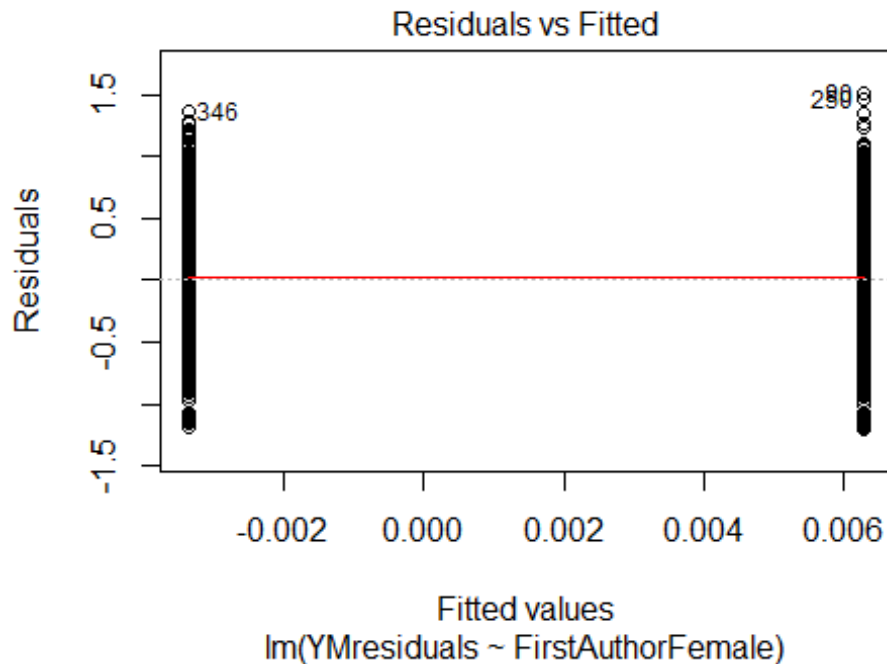
## Year2007      0.08727      0.08183      1.07      0.286
## Year2008      0.08384      0.08035      1.04      0.297
## Year2009      0.05752      0.08059      0.71      0.475
## Year2010      0.14229      0.07967      1.79      0.074 .
## Year2011      0.13883      0.07955      1.75      0.081 .
## Year2012      0.13623      0.07947      1.71      0.087 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0161, Adjusted R-squared:  0.00855
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 2069 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.180  0.873  0.949  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      4.45e-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: 2247"
## [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
## 142 118 102 103 134 182 155 136 168 186 241 242 278 310 282
## 2011 2012
## 343 366
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 114 99 81 87 99 134 137 120 141 162 219 199 241 270 241
## 2011 2012

```

```
## 303 314
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 110 94 75 83 98 127 129 108 128 143 205 184 219 252 228
## 2011 2012
## 287 291
## [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 = 8e-05
```

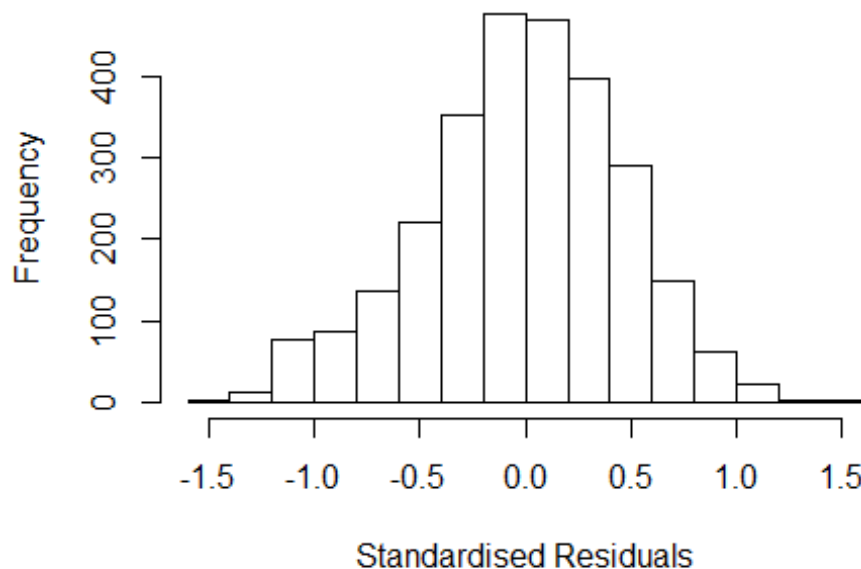


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.04460926203846"
## [1] "Male first author team size 2018 geometric mean: 2.56680102500818"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.00900850868585"
## [1] "Male last author team size 2018 geometric mean: 2.74033290984836"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, 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 1.213 1 1.102
## LastAuthorFemale 1.149 1 1.072
## UniqueAuthors 1.224 4 1.026
## Year 1.279 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
## -1.42330 -0.30097  0.00755  0.31776  1.56831
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03369    0.06231   16.59 < 2e-16 ***
## FirstAuthorFemale1 -0.01463    0.02129   -0.69  0.492
## LastAuthorFemale1 -0.04891    0.01960   -2.50  0.013 *
## UniqueAuthors2     0.14526    0.02629    5.52 3.6e-08 ***
## UniqueAuthors3     0.18129    0.02816    6.44 1.4e-10 ***
## UniqueAuthors4     0.25900    0.03333    7.77 1.1e-14 ***
## UniqueAuthors5     0.32217    0.03399    9.48 < 2e-16 ***
## Year1997         -0.13573    0.08726   -1.56  0.120
## Year1998         -0.03436    0.09023   -0.38  0.703
## Year1999          0.02695    0.09023    0.30  0.765
```

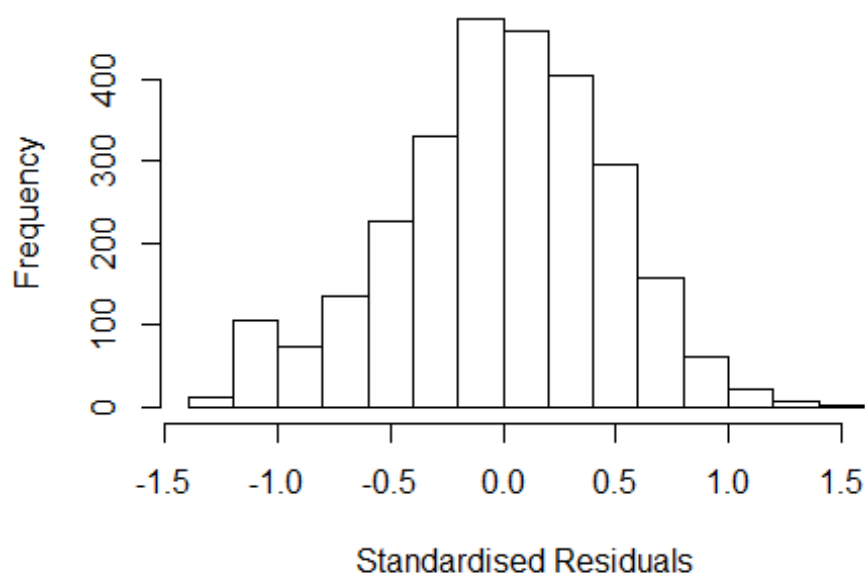


```

## Year2000      -0.00140    0.07892   -0.02    0.986
## Year2001      -0.04319    0.07722   -0.56    0.576
## Year2002       0.00744    0.07441    0.10    0.920
## Year2003       0.12789    0.07201    1.78    0.076 .
## Year2004       0.05205    0.07123    0.73    0.465
## Year2005      -0.02051    0.07224   -0.28    0.776
## Year2006       0.02225    0.06959    0.32    0.749
## Year2007       0.06743    0.06822    0.99    0.323
## Year2008      -0.06409    0.06853   -0.94    0.350
## Year2009      -0.03066    0.06670   -0.46    0.646
## Year2010      -0.00899    0.06847   -0.13    0.896
## Year2011       0.03265    0.06767    0.48    0.630
## Year2012      -0.02610    0.06825   -0.38    0.702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0558, Adjusted R-squared:  0.0482
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 208 weights are ~= 1. The remaining 2553 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.221  0.872  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          3.62e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.190 1          1.091
## LastAuthorFemale  1.156 1          1.075
## Year              1.073 16          1.002

```

Residuals from first and last author



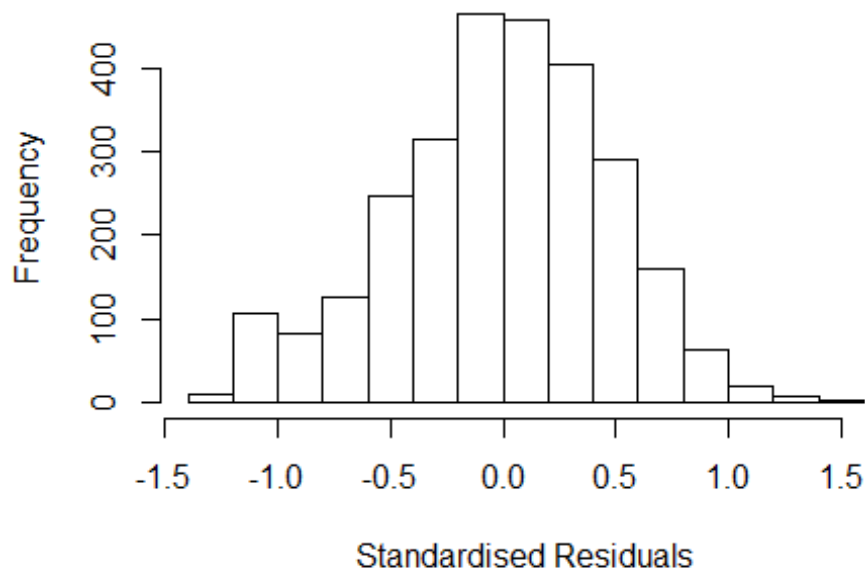
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2436 -0.3146  0.0106  0.3209  1.4805
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12153    0.06168   18.18  <2e-16 ***
## FirstAuthorFemale1  0.01880    0.02166    0.87   0.3854
## LastAuthorFemale1 -0.05575    0.02016   -2.77   0.0057 **
## Year1997        -0.13853    0.08906   -1.56   0.1200
## Year1998        -0.02468    0.09059   -0.27   0.7853
## Year1999         0.03812    0.08989    0.42   0.6715
## Year2000         0.03506    0.07944    0.44   0.6590
## Year2001        -0.01652    0.07939   -0.21   0.8351
## Year2002         0.01639    0.07625    0.21   0.8298
## Year2003         0.16476    0.07277    2.26   0.0237 *
## Year2004         0.09507    0.07153    1.33   0.1840
## Year2005         0.00384    0.07354    0.05   0.9584
```

```

## Year2006          0.06676      0.06988      0.96      0.3395
## Year2007          0.10324      0.06909      1.49      0.1352
## Year2008         -0.02479      0.06938     -0.36      0.7209
## Year2009          0.00760      0.06721      0.11      0.9100
## Year2010          0.03236      0.06917      0.47      0.6399
## Year2011          0.08002      0.06784      1.18      0.2383
## Year2012          0.04179      0.06853      0.61      0.5420
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.47
## Multiple R-squared:  0.0164, Adjusted R-squared:  0.00992
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 238 weights are ~= 1. The remaining 2523 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.299  0.869   0.949   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      3.62e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



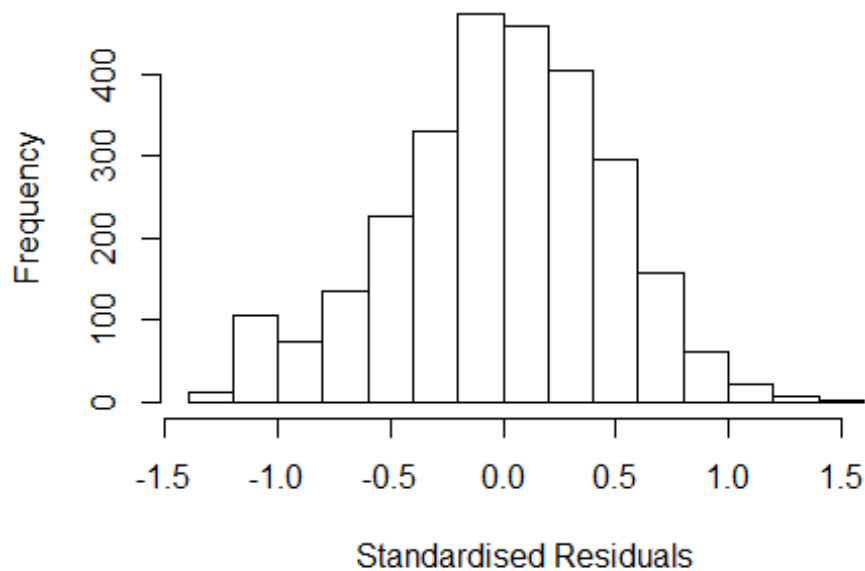
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.20953 -0.32159  0.00962  0.31837  1.49746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.10454    0.06161   17.93  <2e-16 ***
## FirstAuthorFemale1 -0.00295    0.02040   -0.14   0.885
## Year1997         -0.13624    0.08940   -1.52   0.128
## Year1998         -0.01874    0.09115   -0.21   0.837
## Year1999          0.03766    0.09083    0.41   0.678
## Year2000          0.03998    0.07911    0.51   0.613
## Year2001         -0.01389    0.07979   -0.17   0.862
## Year2002          0.01475    0.07628    0.19   0.847
## Year2003          0.17016    0.07253    2.35   0.019 *
## Year2004          0.09979    0.07156    1.39   0.163
## Year2005          0.00493    0.07403    0.07   0.947
## Year2006          0.06664    0.07000    0.95   0.341
```

```

## Year2007          0.10500    0.06928    1.52    0.130
## Year2008         -0.02515    0.06958   -0.36    0.718
## Year2009          0.00773    0.06750    0.11    0.909
## Year2010          0.03112    0.06939    0.45    0.654
## Year2011          0.07904    0.06797    1.16    0.245
## Year2012          0.03904    0.06885    0.57    0.571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00743
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 238 weights are ~= 1. The remaining 2523 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.288  0.869  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      3.62e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



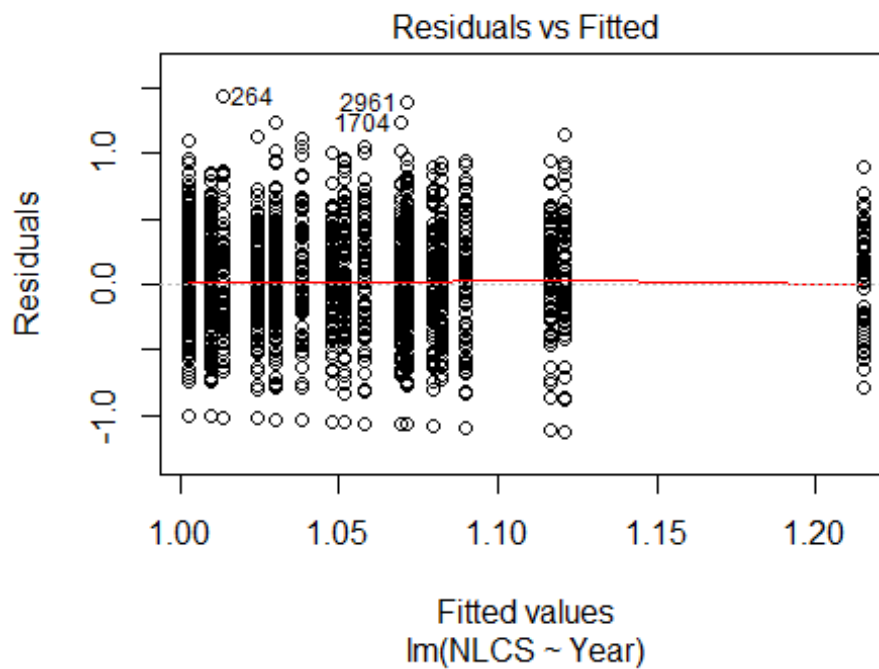
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2345 -0.3147 0.0113 0.3242 1.4748
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12721 0.06162 18.29 <2e-16 ***
## LastAuthorFemale1 -0.04914 0.01904 -2.58 0.0099 **
## Year1997 -0.13726 0.08933 -1.54 0.1245
## Year1998 -0.02293 0.09097 -0.25 0.8010
## Year1999 0.03983 0.08984 0.44 0.6576
## Year2000 0.03677 0.07952 0.46 0.6438
## Year2001 -0.01390 0.07947 -0.17 0.8612
## Year2002 0.01794 0.07649 0.23 0.8146
## Year2003 0.16825 0.07259 2.32 0.0205 *
## Year2004 0.09766 0.07157 1.36 0.1725
## Year2005 0.00668 0.07357 0.09 0.9277
## Year2006 0.06958 0.06979 1.00 0.3188
```

```

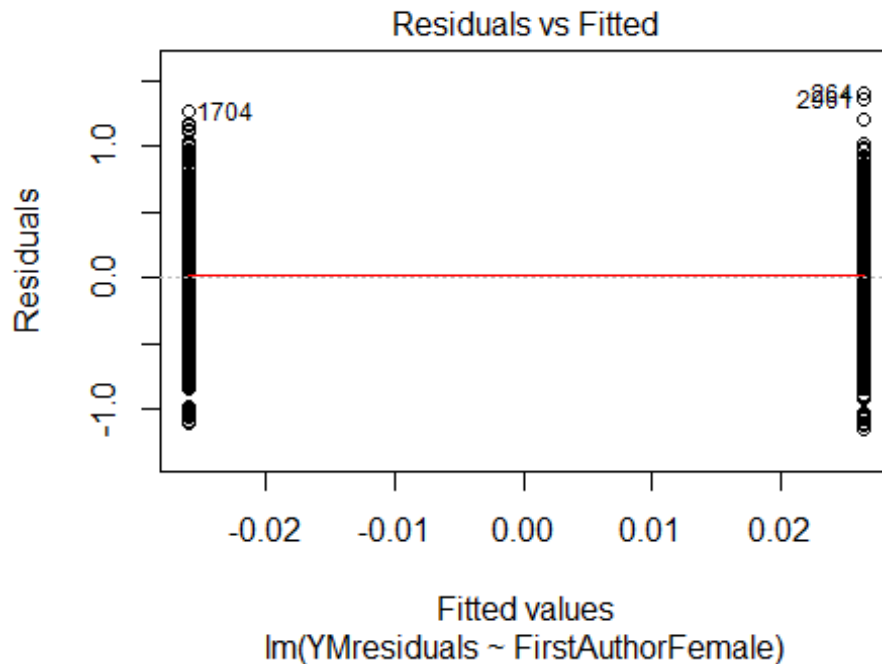
## Year2007      0.10730      0.06902      1.55      0.1201
## Year2008     -0.02204      0.06935     -0.32      0.7506
## Year2009      0.01152      0.06713      0.17      0.8637
## Year2010      0.03649      0.06899      0.53      0.5969
## Year2011      0.08389      0.06764      1.24      0.2150
## Year2012      0.04571      0.06839      0.67      0.5040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.47
## Multiple R-squared:  0.0162, Adjusted R-squared:  0.0101
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 244 weights are ~= 1. The remaining 2517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.303  0.867  0.949  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      3.62e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2761"
## [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
##  104  102   86   97   94  108  137  106  130  134  144  137  187  178  207
## 2011 2012
##  231  211
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   86   81   69   74   69   77  115   88  103  113  126  112  163  161  173
## 2011 2012

```

```
## 206 178
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 76 64 63 65 68 103 78 92 98 114 105 147 157 158
## 2011 2012
## 198 167
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 40, df = 16, p-value = 7e-04
```

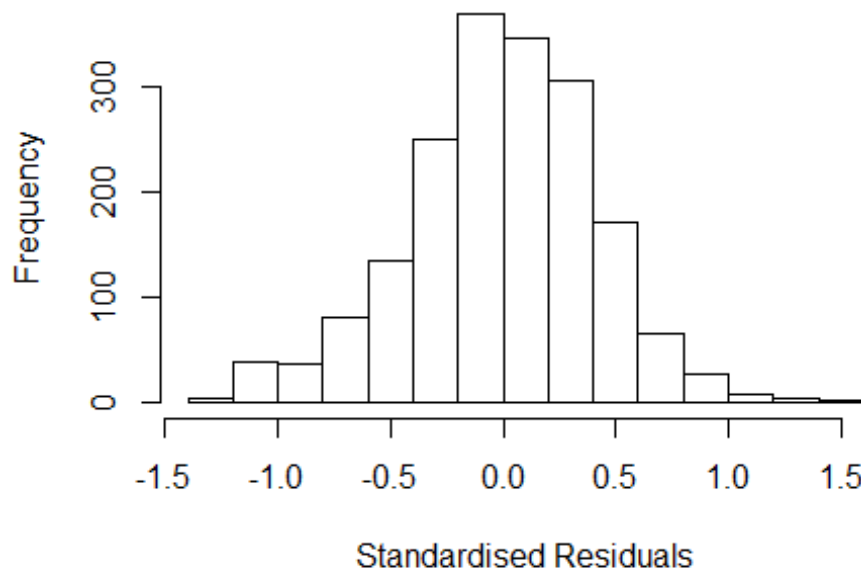


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.36, df = 1, p-value = 0.5
```

```
## [1] "Female first author team size 2018 geometric mean: 3.12925841351139"
## [1] "Male first author team size 2018 geometric mean: 2.94765461455369"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6800, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.90425032350332"
## [1] "Male last author team size 2018 geometric mean: 3.16166936357572"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5900, 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 1.140 1          1.068
## LastAuthorFemale  1.087 1          1.043
## UniqueAuthors    1.226 4          1.026
## Year             1.373 16          1.010
```

Residuals from first and last author and team size



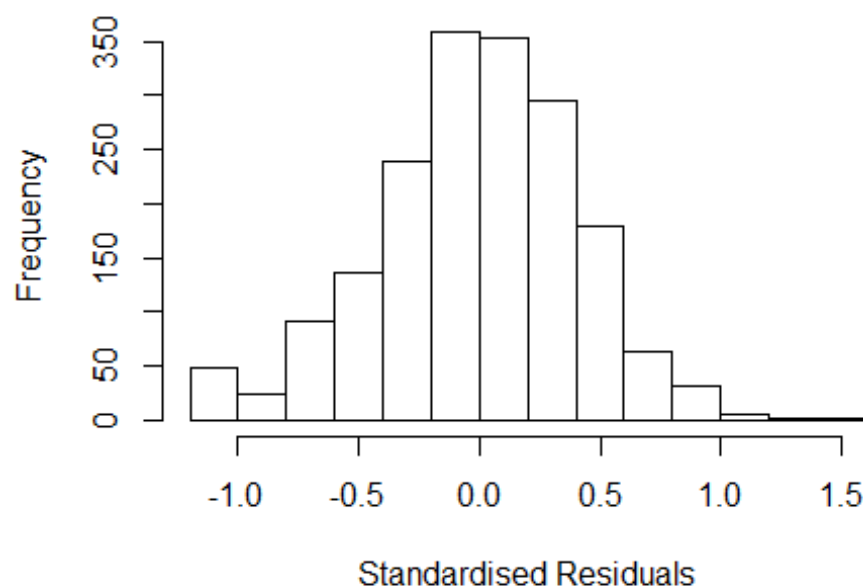
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22876 -0.26795  0.00559  0.26244  1.52121
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0185    0.0640   15.92 < 2e-16 ***
## FirstAuthorFemale1  0.0409    0.0204    2.00  0.04524 *
## LastAuthorFemale1  0.0175    0.0202    0.87  0.38631
## UniqueAuthors2     0.1323    0.0336    3.94  8.5e-05 ***
## UniqueAuthors3     0.1226    0.0349    3.51  0.00046 ***
## UniqueAuthors4     0.2139    0.0411    5.20  2.2e-07 ***
## UniqueAuthors5     0.2301    0.0403    5.70  1.4e-08 ***
## Year1997          -0.1008    0.0829   -1.22  0.22391
## Year1998          -0.1542    0.0832   -1.85  0.06392 .
## Year1999          -0.0738    0.0860   -0.86  0.39067
```

```

## Year2000          0.0490      0.0741      0.66  0.50876
## Year2001         -0.0266      0.0805     -0.33  0.74141
## Year2002         -0.0608      0.0705     -0.86  0.38852
## Year2003         -0.0433      0.0785     -0.55  0.58143
## Year2004         -0.0819      0.0728     -1.13  0.26054
## Year2005         -0.0935      0.0678     -1.38  0.16841
## Year2006         -0.1180      0.0682     -1.73  0.08371 .
## Year2007         -0.1205      0.0680     -1.77  0.07659 .
## Year2008         -0.1021      0.0657     -1.55  0.12074
## Year2009         -0.1656      0.0658     -2.52  0.01192 *
## Year2010         -0.1769      0.0675     -2.62  0.00889 **
## Year2011         -0.1542      0.0660     -2.34  0.01965 *
## Year2012         -0.0917      0.0680     -1.35  0.17762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.0449, Adjusted R-squared:  0.0333
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1678 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.107  0.869  0.951  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          5.45e-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 1.114 1          1.055
## LastAuthorFemale 1.073 1          1.036
## Year              1.131 16          1.004

```

Residuals from first and last author



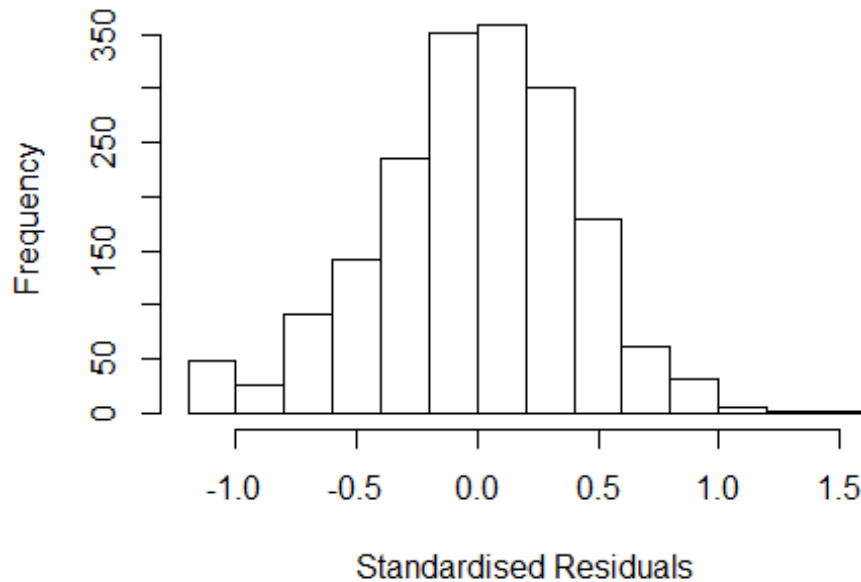
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19370 -0.27022 0.00815 0.26633 1.40298
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.110852 0.059168 18.77 <2e-16 ***
## FirstAuthorFemale1 0.056157 0.020568 2.73 0.0064 **
## LastAuthorFemale1 0.018608 0.020503 0.91 0.3642
## Year1997 -0.082320 0.084206 -0.98 0.3284
## Year1998 -0.144596 0.081829 -1.77 0.0774 .
## Year1999 -0.067408 0.087793 -0.77 0.4427
## Year2000 0.059931 0.073008 0.82 0.4118
## Year2001 0.008085 0.079782 0.10 0.9193
## Year2002 -0.050081 0.072415 -0.69 0.4893
## Year2003 0.000141 0.078437 0.00 0.9986
## Year2004 -0.044643 0.073346 -0.61 0.5428
## Year2005 -0.060690 0.068274 -0.89 0.3742
```

```

## Year2006      -0.100571    0.068218    -1.47    0.1406
## Year2007      -0.086055    0.068617    -1.25    0.2100
## Year2008      -0.066093    0.066404    -1.00    0.3197
## Year2009      -0.126382    0.066381    -1.90    0.0571 .
## Year2010      -0.137879    0.068339    -2.02    0.0438 *
## Year2011      -0.131768    0.066866    -1.97    0.0489 *
## Year2012      -0.048837    0.068559    -0.71    0.4763
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0195, Adjusted R-squared:  0.00982
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1686 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  0.869  0.950  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      5.45e-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 1.085 1      1.042
## Year      1.085 16      1.003

```

Residuals from first author



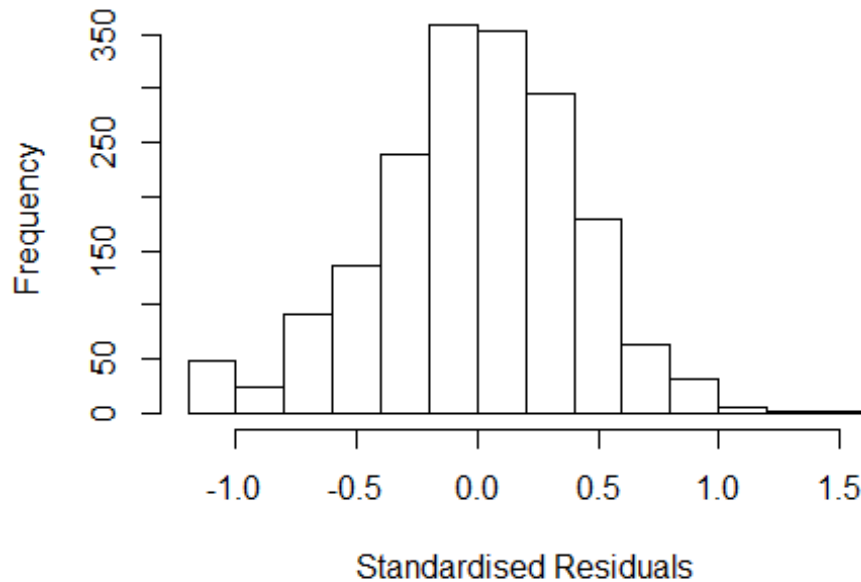
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1832 -0.2713 0.0104 0.2633 1.4132
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.116322 0.058644 19.04 <2e-16 ***
## FirstAuthorFemale1 0.060154 0.020332 2.96 0.0031 **
## Year1997 -0.082343 0.084193 -0.98 0.3282
## Year1998 -0.145708 0.081698 -1.78 0.0747 .
## Year1999 -0.069347 0.087686 -0.79 0.4291
## Year2000 0.060017 0.072988 0.82 0.4110
## Year2001 0.006674 0.079506 0.08 0.9331
## Year2002 -0.050030 0.072242 -0.69 0.4887
## Year2003 0.000267 0.078223 0.00 0.9973
## Year2004 -0.042765 0.072934 -0.59 0.5577
## Year2005 -0.061313 0.068045 -0.90 0.3677
## Year2006 -0.099531 0.067827 -1.47 0.1424
```

```

## Year2007          -0.088087    0.068432   -1.29    0.1982
## Year2008          -0.065364    0.066139   -0.99    0.3231
## Year2009          -0.126180    0.066128   -1.91    0.0565 .
## Year2010          -0.137741    0.068010   -2.03    0.0430 *
## Year2011          -0.131567    0.066586   -1.98    0.0483 *
## Year2012          -0.050016    0.068217   -0.73    0.4635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.019, Adjusted R-squared:  0.00987
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1686 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.186  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      5.45e-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 1.046 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
## -1.16888 -0.27256 0.00922 0.26947 1.42660
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12602 0.05917 19.03 <2e-16 ***
## LastAuthorFemale1 0.03093 0.02030 1.52 0.128
## Year1997 -0.08345 0.08417 -0.99 0.322
## Year1998 -0.13954 0.08190 -1.70 0.089 .
## Year1999 -0.05717 0.08879 -0.64 0.520
## Year2000 0.05820 0.07301 0.80 0.425
## Year2001 0.01194 0.08051 0.15 0.882
## Year2002 -0.04691 0.07271 -0.65 0.519
## Year2003 0.00893 0.07879 0.11 0.910
## Year2004 -0.04526 0.07382 -0.61 0.540
## Year2005 -0.05238 0.06864 -0.76 0.446
## Year2006 -0.09279 0.06827 -1.36 0.174
```

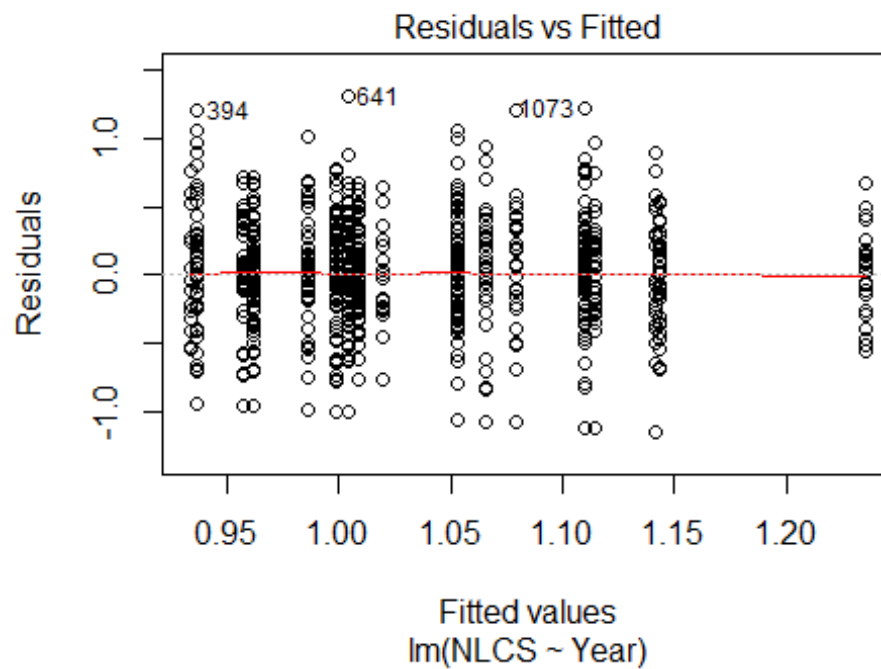


```

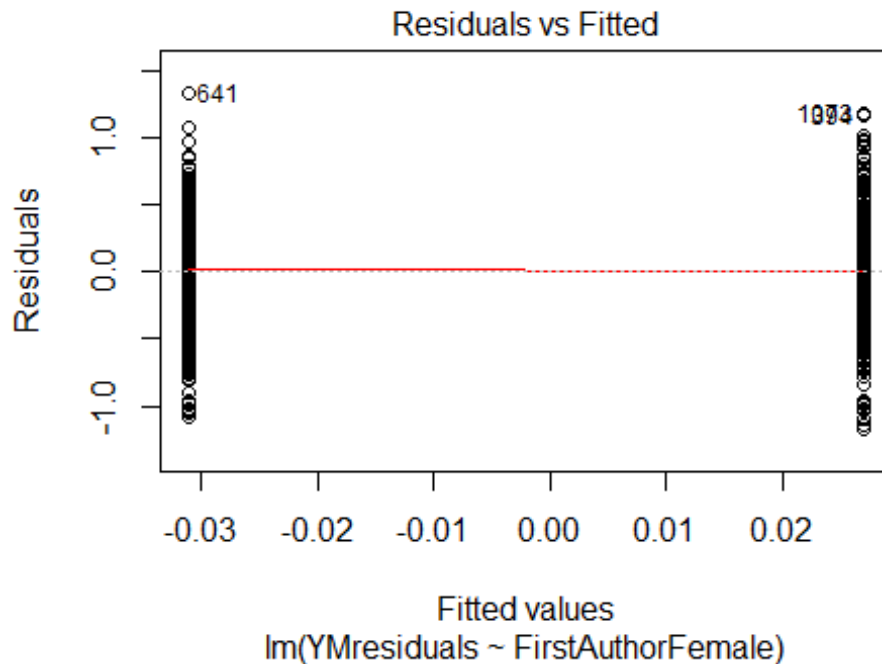
## Year2007          -0.07521      0.06876    -1.09      0.274
## Year2008          -0.05651      0.06634    -0.85      0.394
## Year2009          -0.11512      0.06629    -1.74      0.083 .
## Year2010          -0.12183      0.06813    -1.79      0.074 .
## Year2011          -0.12257      0.06672    -1.84      0.066 .
## Year2012          -0.03461      0.06827    -0.51      0.612
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.00598
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 155 weights are ~= 1. The remaining 1680 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.179  0.873  0.951  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      5.45e-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: 1835"
## [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
##   29   19   41   21   49   50   52   40   62   46   77   54   92   98   72
## 2011 2012
##   97   85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   14   32   20   27   29   45   34   49   39   67   43   82   91   63
## 2011 2012

```

```
##      86      76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   11   29   16   25   24   42   30   46   36   61   41   74   87   58
## 2011 2012
##   80   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 = 51, df = 16, p-value = 2e-05
```

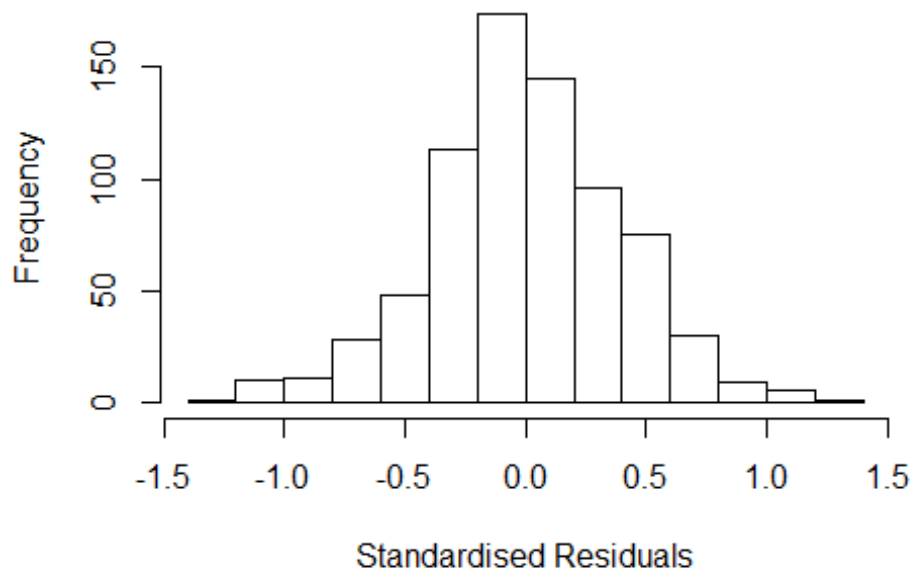


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.71, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.7002500004948"
## [1] "Male first author team size 2018 geometric mean: 3.0508018829927"
##
## 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: 3.49574870413595"
## [1] "Male last author team size 2018 geometric mean: 3.36303166037009"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, 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 1.178 1      1.086
## LastAuthorFemale  1.164 1      1.079
## UniqueAuthors    1.739 4      1.072
## Year              1.903 16     1.020
```

Residuals from first and last author and team size



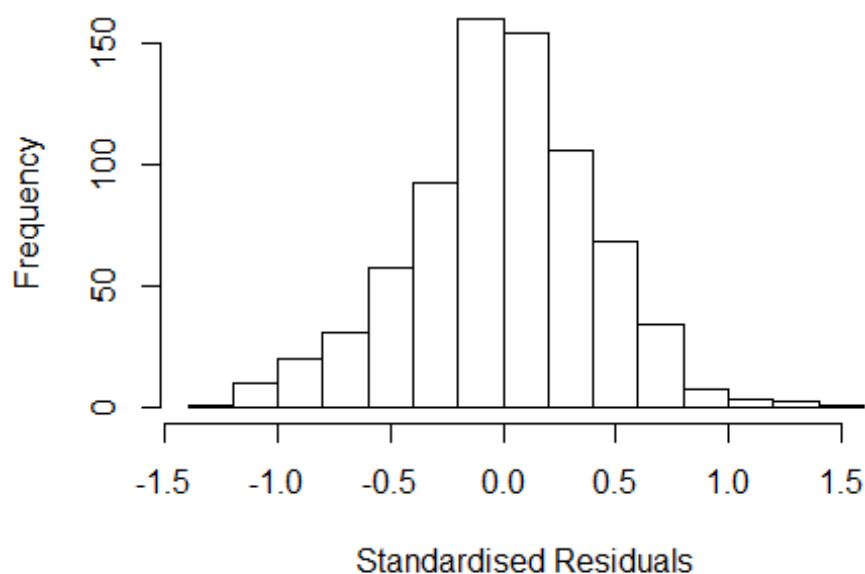
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3132 -0.2386 -0.0149 0.2439 1.3448
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8563 0.1616 5.30 1.5e-07 ***
## FirstAuthorFemale1 0.0197 0.0304 0.65 0.517
## LastAuthorFemale1 0.0451 0.0300 1.50 0.133
## UniqueAuthors2 0.3921 0.0610 6.43 2.3e-10 ***
## UniqueAuthors3 0.3187 0.0592 5.38 9.8e-08 ***
## UniqueAuthors4 0.4233 0.0615 6.88 1.3e-11 ***
## UniqueAuthors5 0.3996 0.0619 6.45 2.0e-10 ***
## Year1997 -0.1569 0.1975 -0.79 0.427
## Year1998 -0.0407 0.1692 -0.24 0.810
## Year1999 -0.1241 0.1827 -0.68 0.497
```

```

## Year2000          -0.0238      0.1657   -0.14    0.886
## Year2001          -0.1570      0.1789   -0.88    0.380
## Year2002          -0.1857      0.1650   -1.13    0.261
## Year2003          -0.2447      0.1707   -1.43    0.152
## Year2004          -0.2901      0.1703   -1.70    0.089 .
## Year2005          -0.1254      0.1735   -0.72    0.470
## Year2006          -0.1521      0.1625   -0.94    0.350
## Year2007          -0.0833      0.1598   -0.52    0.602
## Year2008          -0.2853      0.1610   -1.77    0.077 .
## Year2009          -0.2554      0.1571   -1.63    0.104
## Year2010          -0.2741      0.1636   -1.68    0.094 .
## Year2011          -0.1747      0.1591   -1.10    0.273
## Year2012          -0.1449      0.1624   -0.89    0.373
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.368
## Multiple R-squared:  0.144, Adjusted R-squared:  0.118
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 678 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.154  0.862  0.949   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      1.34e-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 1.084 1      1.041
## LastAuthorFemale  1.122 1      1.059
## Year              1.216 16      1.006

```

Residuals from first and last author



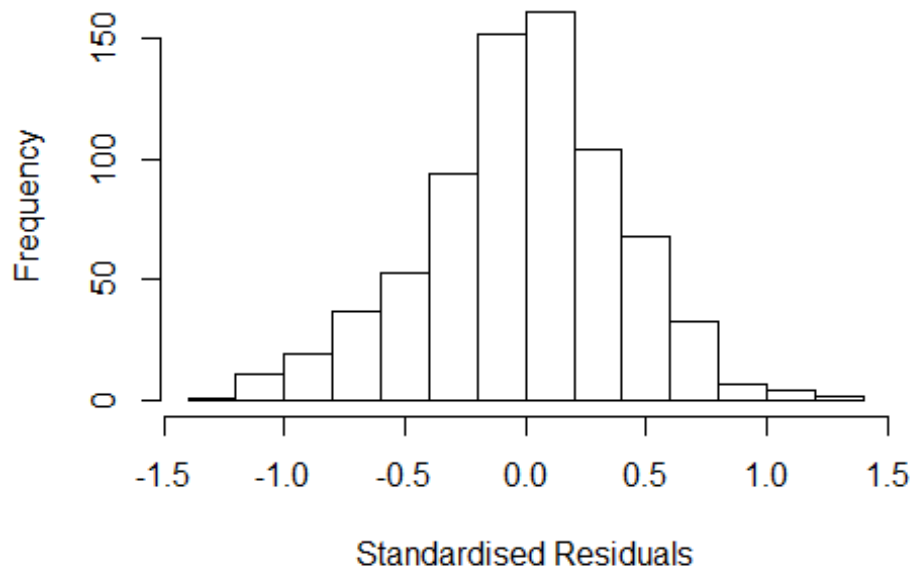
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24782 -0.24971  0.00236  0.25289  1.41205
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1353     0.1572   7.22 1.3e-12 ***
## FirstAuthorFemale1  0.0569     0.0311   1.83  0.068 .
## LastAuthorFemale1  0.0557     0.0312   1.78  0.075 .
## Year1997         -0.1673     0.2105  -0.79  0.427
## Year1998         -0.0707     0.1716  -0.41  0.680
## Year1999         -0.1620     0.1842  -0.88  0.379
## Year2000          0.0236     0.1679   0.14  0.888
## Year2001         -0.1391     0.1890  -0.74  0.462
## Year2002         -0.1975     0.1717  -1.15  0.250
## Year2003         -0.2155     0.1775  -1.21  0.225
## Year2004         -0.2688     0.1826  -1.47  0.141
## Year2005         -0.0922     0.1789  -0.52  0.606
```

```

## Year2006          -0.1515      0.1682   -0.90    0.368
## Year2007          -0.0434      0.1649   -0.26    0.793
## Year2008          -0.2393      0.1643   -1.46    0.146
## Year2009          -0.1919      0.1600   -1.20    0.231
## Year2010          -0.2522      0.1667   -1.51    0.131
## Year2011          -0.1645      0.1625   -1.01    0.312
## Year2012          -0.0861      0.1662   -0.52    0.605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0467, Adjusted R-squared:  0.0231
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 671 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.134  0.855  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      1.34e-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 1.082 1          1.040
## Year              1.082 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
## -1.23199 -0.23597  0.00919  0.24797  1.38801
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1660    0.1553   7.51 1.7e-13 ***
## FirstAuthorFemale1  0.0660    0.0316   2.09  0.037 *
## Year1997         -0.1691    0.2133  -0.79  0.428
## Year1998         -0.0778    0.1712  -0.45  0.650
## Year1999         -0.1694    0.1827  -0.93  0.354
## Year2000          0.0110    0.1669   0.07  0.948
## Year2001         -0.1634    0.1882  -0.87  0.385
## Year2002         -0.2114    0.1719  -1.23  0.219
## Year2003         -0.2375    0.1765  -1.35  0.179
## Year2004         -0.2772    0.1818  -1.53  0.128
## Year2005         -0.0929    0.1775  -0.52  0.601
## Year2006         -0.1659    0.1671  -0.99  0.321
```

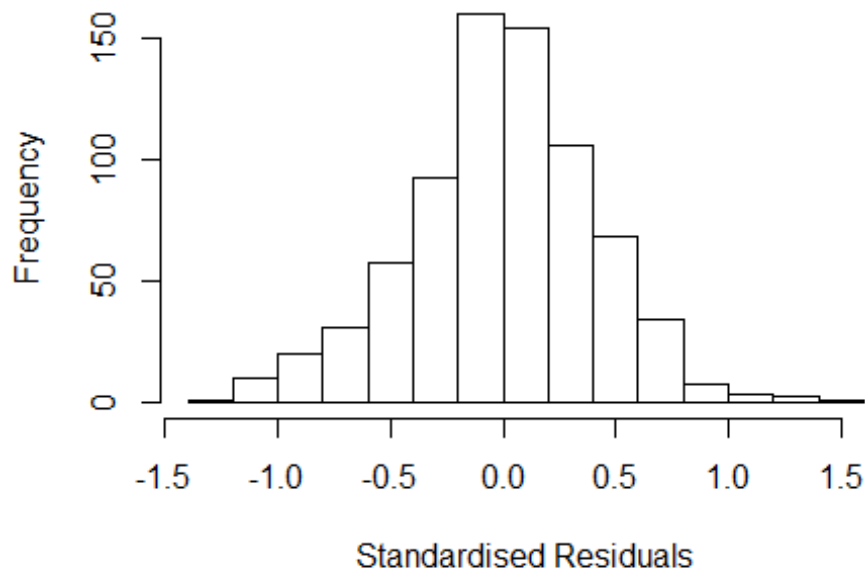


```

## Year2007          -0.0633      0.1638   -0.39    0.699
## Year2008          -0.2460      0.1635   -1.50    0.133
## Year2009          -0.2043      0.1590   -1.28    0.199
## Year2010          -0.2614      0.1656   -1.58    0.115
## Year2011          -0.1808      0.1614   -1.12    0.263
## Year2012          -0.0919      0.1648   -0.56    0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0423, Adjusted R-squared:  0.0199
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 681 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.852  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      1.34e-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 1.123 1      1.060
## Year              1.123 16      1.004

```

Residuals from last author



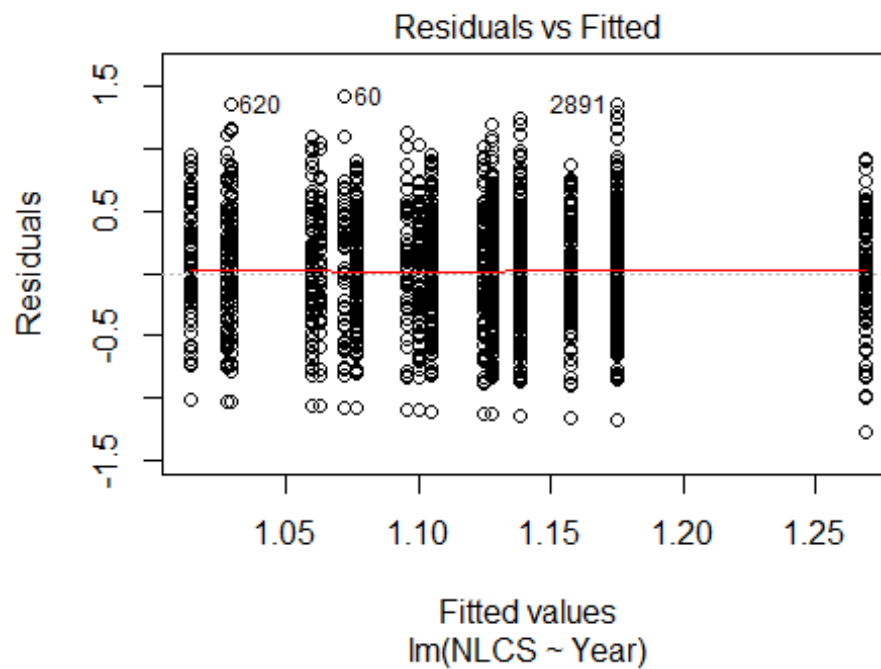
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.220575 -0.249268  0.000431  0.242789  1.382283
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1556     0.1557   7.42 3.2e-13 ***
## LastAuthorFemale1  0.0650     0.0316   2.06  0.04 *
## Year1997        -0.1822     0.2094  -0.87  0.38
## Year1998        -0.0651     0.1690  -0.39  0.70
## Year1999        -0.1530     0.1834  -0.83  0.40
## Year2000         0.0268     0.1665   0.16  0.87
## Year2001        -0.1412     0.1886  -0.75  0.45
## Year2002        -0.1915     0.1700  -1.13  0.26
## Year2003        -0.2136     0.1760  -1.21  0.23
## Year2004        -0.2733     0.1827  -1.50  0.14
## Year2005        -0.0806     0.1776  -0.45  0.65
## Year2006        -0.1493     0.1669  -0.89  0.37
```

```

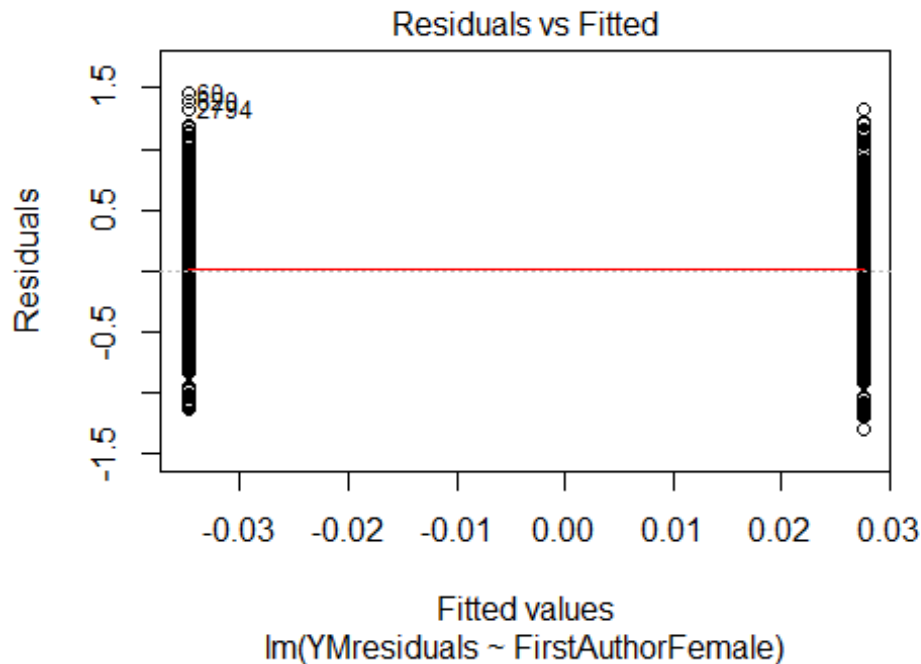
## Year2007          -0.0317      0.1626   -0.20      0.85
## Year2008          -0.2299      0.1619   -1.42      0.16
## Year2009          -0.1822      0.1577   -1.16      0.25
## Year2010          -0.2358      0.1641   -1.44      0.15
## Year2011          -0.1545      0.1603   -0.96      0.34
## Year2012          -0.0779      0.1641   -0.47      0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0421, Adjusted R-squared:  0.0198
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 687 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.152  0.857  0.950  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      1.34e-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: 746"
## [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
##   69   67   85   90  123  127  133   79   92  109  116  128  191  224  224
## 2011 2012
##   224  238
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   55   55   60   72   85   94  102   65   84  106   95  108  171  201  189
## 2011 2012

```

```
## 196 200
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 50 53 66 80 85 94 59 81 98 88 100 158 189 178
## 2011 2012
## 180 182
## [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
```

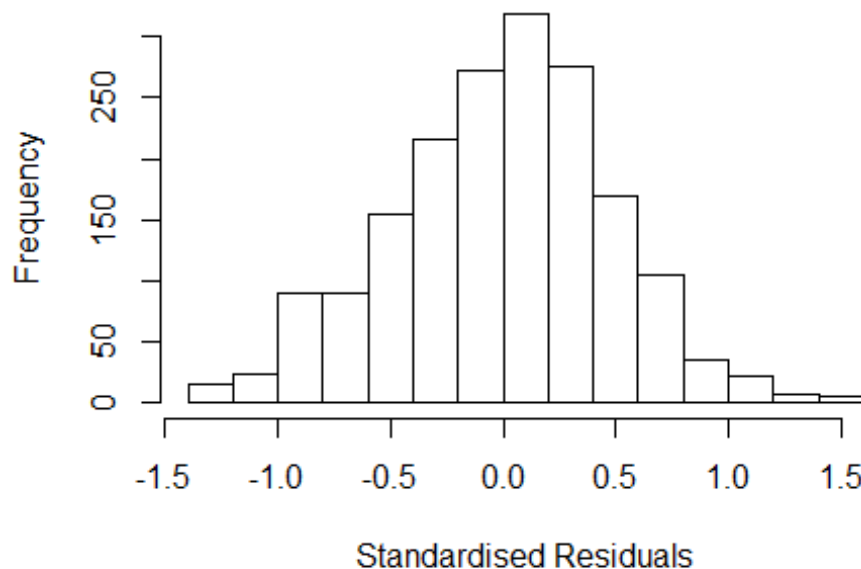


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.8, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.42583592197322"
## [1] "Male first author team size 2018 geometric mean: 2.13187372593058"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7300, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.37657919436351"
## [1] "Male last author team size 2018 geometric mean: 2.23554457597302"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7200, 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 1.224 1          1.106
## LastAuthorFemale  1.176 1          1.084
## UniqueAuthors    1.260 4          1.029
## Year              1.325 16         1.009
```

Residuals from first and last author and team size



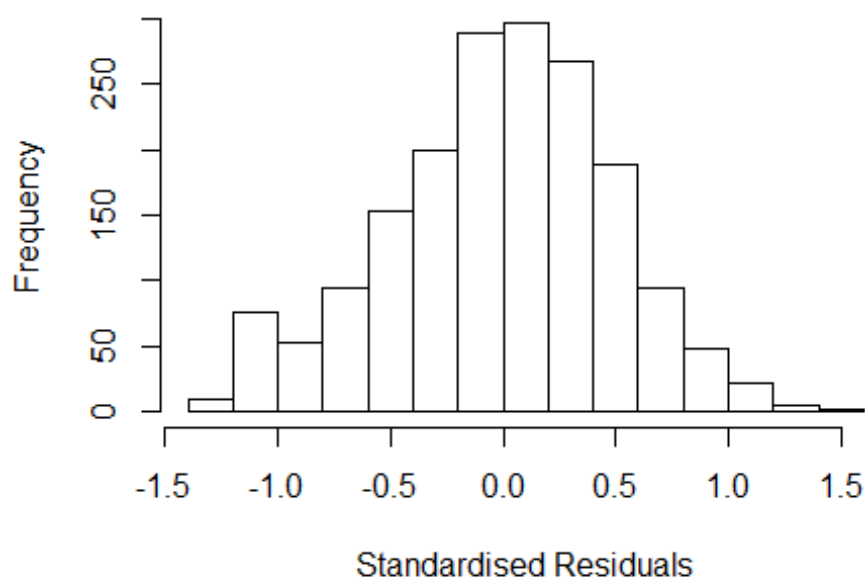
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3504 -0.3134 0.0239 0.3188 1.5077
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98428 0.08302 11.86 < 2e-16 ***
## FirstAuthorFemale1 0.06299 0.02585 2.44 0.015 *
## LastAuthorFemale1 -0.05123 0.02500 -2.05 0.041 *
## UniqueAuthors2 0.22599 0.03049 7.41 1.9e-13 ***
## UniqueAuthors3 0.24425 0.03421 7.14 1.4e-12 ***
## UniqueAuthors4 0.24739 0.04506 5.49 4.6e-08 ***
## UniqueAuthors5 0.32311 0.04541 7.12 1.6e-12 ***
## Year1997 -0.03398 0.11012 -0.31 0.758
## Year1998 -0.02146 0.11384 -0.19 0.851
## Year1999 -0.12518 0.10529 -1.19 0.235
```

```

## Year2000      -0.04625    0.09898   -0.47    0.640
## Year2001      -0.03614    0.10143   -0.36    0.722
## Year2002      -0.07252    0.09675   -0.75    0.454
## Year2003       0.12838    0.10798    1.19    0.235
## Year2004       0.12254    0.09668    1.27    0.205
## Year2005      -0.02795    0.09091   -0.31    0.759
## Year2006      -0.01623    0.09588   -0.17    0.866
## Year2007      -0.00336    0.09008   -0.04    0.970
## Year2008      -0.07934    0.08783   -0.90    0.366
## Year2009      -0.04242    0.08589   -0.49    0.621
## Year2010      -0.02445    0.09001   -0.27    0.786
## Year2011      -0.02592    0.08928   -0.29    0.772
## Year2012      -0.01328    0.08724   -0.15    0.879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.0745, Adjusted R-squared:  0.063
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~= 1. The remaining 1633 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.278  0.864  0.949  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      5.57e-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 1.207 1      1.099
## LastAuthorFemale  1.139 1      1.067
## 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
## -1.3039 -0.3230 0.0186 0.3339 1.4241
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06786 0.08739 12.22 <2e-16 ***
## FirstAuthorFemale1 0.08693 0.02649 3.28 0.0011 **
## LastAuthorFemale1 -0.05415 0.02552 -2.12 0.0340 *
## Year1997 -0.03192 0.11381 -0.28 0.7791
## Year1998 0.02262 0.11615 0.19 0.8456
## Year1999 -0.09981 0.10667 -0.94 0.3495
## Year2000 -0.03046 0.10255 -0.30 0.7665
## Year2001 0.01080 0.10812 0.10 0.9204
## Year2002 -0.02536 0.10173 -0.25 0.8032
## Year2003 0.16459 0.11451 1.44 0.1508
## Year2004 0.20322 0.09942 2.04 0.0411 *
## Year2005 0.03212 0.09560 0.34 0.7369
```

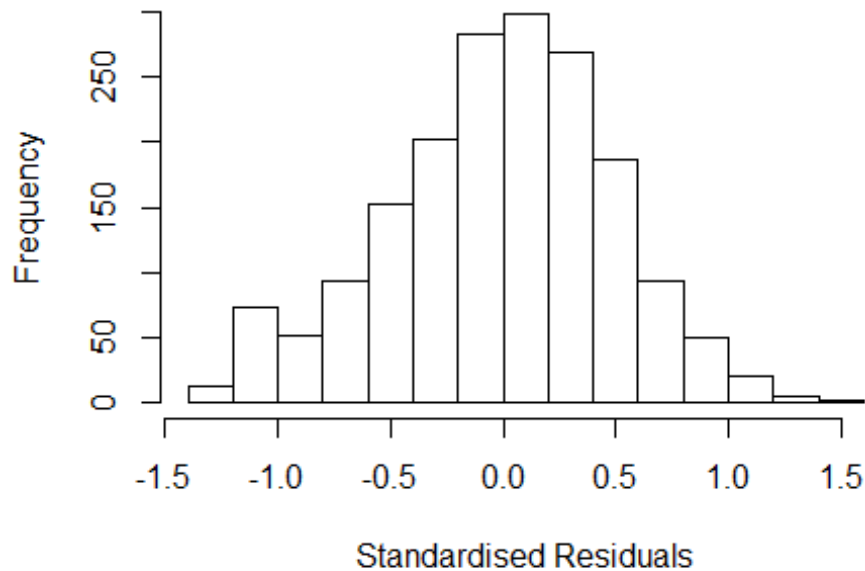


```

## Year2006          0.07105      0.10004      0.71      0.4777
## Year2007          0.09327      0.09473      0.98      0.3250
## Year2008         -0.00578      0.09303     -0.06      0.9505
## Year2009          0.03074      0.09044      0.34      0.7340
## Year2010          0.05004      0.09416      0.53      0.5952
## Year2011          0.05041      0.09396      0.54      0.5917
## Year2012          0.08626      0.09123      0.95      0.3445
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.0218, Adjusted R-squared:  0.0119
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1666 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.370  0.872  0.950  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      5.57e-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 1.084 1      1.041
## Year              1.084 16      1.003

```

Residuals from first author



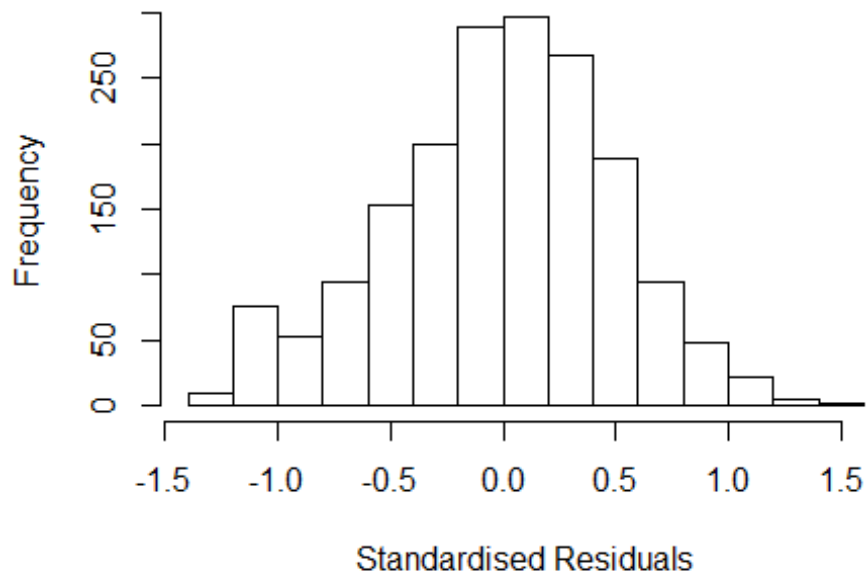
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3184 -0.3280 0.0238 0.3318 1.4423
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04973 0.08572 12.25 <2e-16 ***
## FirstAuthorFemale1 0.06401 0.02523 2.54 0.011 *
## Year1997 -0.02984 0.11332 -0.26 0.792
## Year1998 0.02879 0.11528 0.25 0.803
## Year1999 -0.09348 0.10652 -0.88 0.380
## Year2000 -0.02367 0.10128 -0.23 0.815
## Year2001 0.01388 0.10715 0.13 0.897
## Year2002 -0.02314 0.10077 -0.23 0.818
## Year2003 0.16897 0.11301 1.50 0.135
## Year2004 0.20469 0.09865 2.07 0.038 *
## Year2005 0.03501 0.09467 0.37 0.712
## Year2006 0.07304 0.09929 0.74 0.462
```

```

## Year2007          0.09652    0.09378    1.03    0.304
## Year2008         -0.00555    0.09214   -0.06    0.952
## Year2009          0.03255    0.08947    0.36    0.716
## Year2010          0.05339    0.09323    0.57    0.567
## Year2011          0.05548    0.09308    0.60    0.551
## Year2012          0.09266    0.09026    1.03    0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0195, Adjusted R-squared:  0.0102
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1664 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.361  0.871  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      5.57e-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 1.029 1          1.014
## Year            1.029 16          1.001

```

Residuals from last author



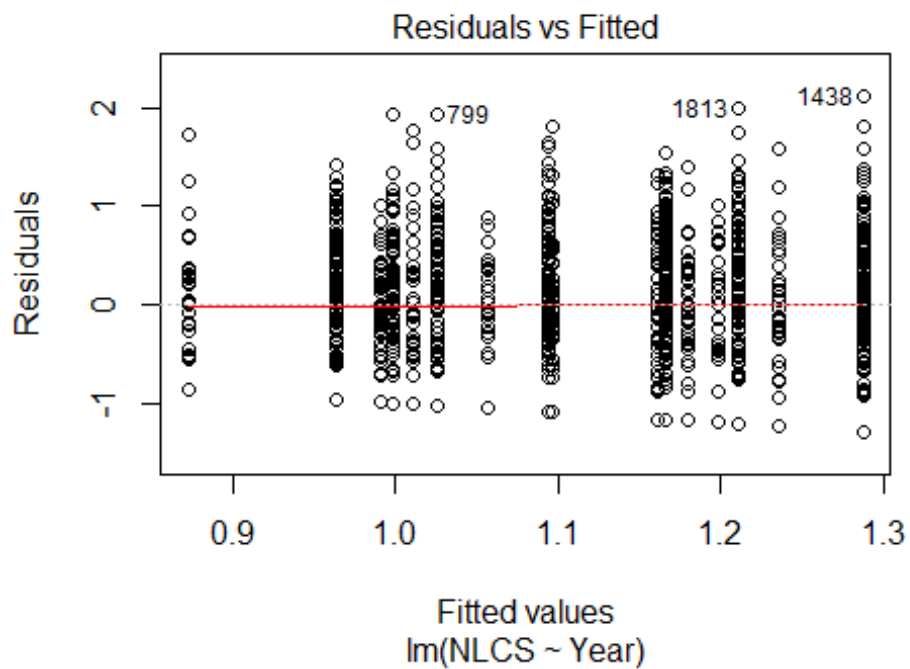
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2739 -0.3230 0.0212 0.3309 1.3920
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10003 0.08785 12.52 <2e-16 ***
## LastAuthorFemale1 -0.01859 0.02445 -0.76 0.447
## Year1997 -0.03134 0.11557 -0.27 0.786
## Year1998 0.01204 0.11702 0.10 0.918
## Year1999 -0.11040 0.10764 -1.03 0.305
## Year2000 -0.04231 0.10413 -0.41 0.685
## Year2001 0.00562 0.10968 0.05 0.959
## Year2002 -0.03189 0.10318 -0.31 0.757
## Year2003 0.15841 0.11477 1.38 0.168
## Year2004 0.19249 0.10087 1.91 0.057 .
## Year2005 0.03090 0.09712 0.32 0.750
## Year2006 0.07559 0.10180 0.74 0.458
```

```

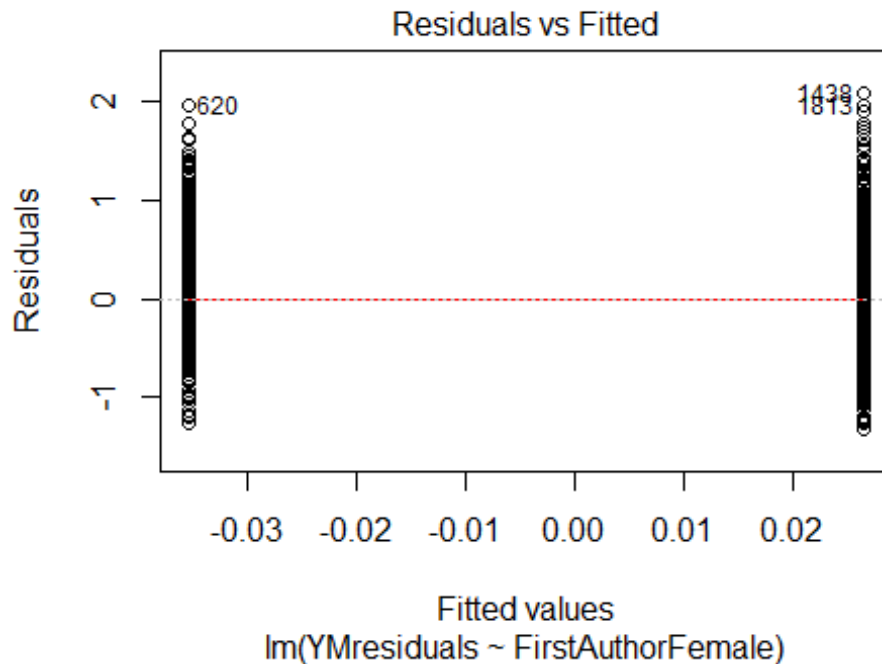
## Year2007      0.09053    0.09663    0.94    0.349
## Year2008     -0.01345    0.09438   -0.14    0.887
## Year2009      0.03249    0.09203    0.35    0.724
## Year2010      0.05233    0.09562    0.55    0.584
## Year2011      0.05639    0.09567    0.59    0.556
## Year2012      0.09403    0.09293    1.01    0.312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0161, Adjusted R-squared:  0.00672
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 135 weights are ~= 1. The remaining 1660 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.396  0.872  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      5.57e-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: 1795"
## [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
##   40   39   64   74   64   48   53   38   58   56  100  131  173  164  146
## 2011 2012
##  194  183
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   28   41   52   46   33   39   30   48   44   88  116  148  144  119
## 2011 2012

```

```
## 168 150
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 27 39 52 42 30 39 28 45 44 83 110 144 135 107
## 2011 2012
## 157 140
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

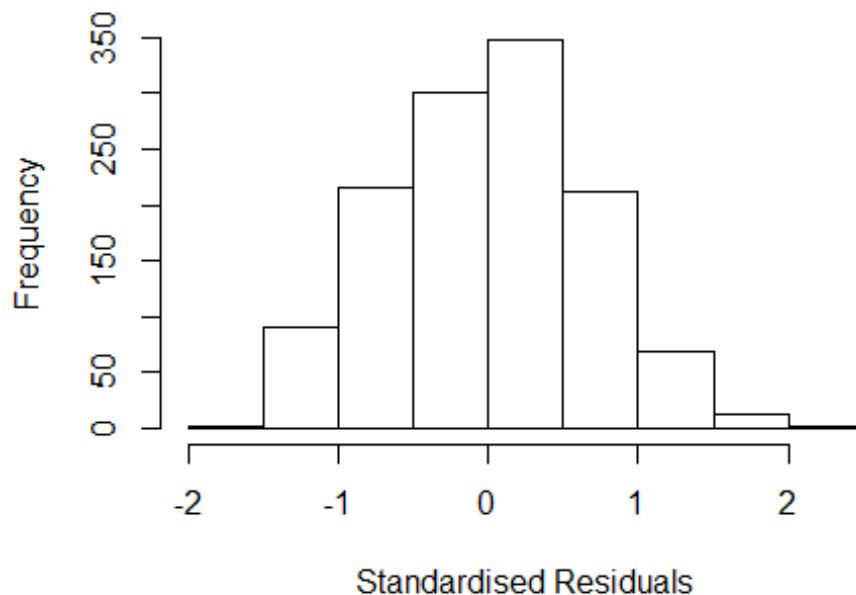


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.89, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.73735968774828"
## [1] "Male first author team size 2018 geometric mean: 1.90615673442136"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3600, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.79714141616494"
## [1] "Male last author team size 2018 geometric mean: 1.8531208256524"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3800, 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 2.093 1      1.447
## LastAuthorFemale  1.938 1      1.392
## UniqueAuthors    1.359 4      1.039
## Year             1.483 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
## -1.5342 -0.4898 0.0179 0.4610 2.0443
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8544 0.1189 7.19 1.2e-12 ***
## FirstAuthorFemale1 -0.0634 0.0559 -1.13 0.25706
## LastAuthorFemale1 -0.0131 0.0555 -0.24 0.81397
## UniqueAuthors2 0.1862 0.0511 3.64 0.00028 ***
## UniqueAuthors3 0.2183 0.0624 3.50 0.00048 ***
## UniqueAuthors4 0.3466 0.1303 2.66 0.00794 **
## UniqueAuthors5 0.2298 0.1129 2.04 0.04206 *
## Year1997 0.1629 0.1526 1.07 0.28620
## Year1998 0.0616 0.1574 0.39 0.69541
## Year1999 0.0853 0.1368 0.62 0.53279
```

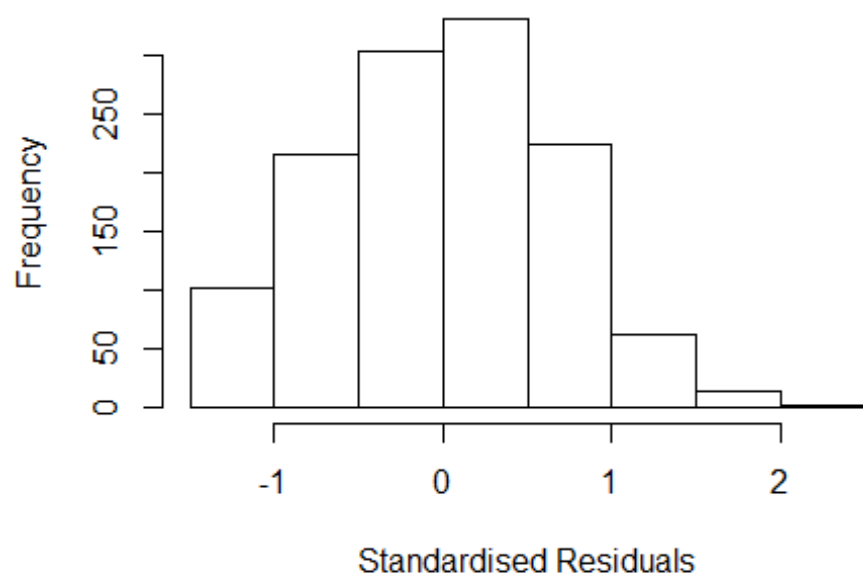


```

## Year2000          0.1603      0.1512      1.06  0.28912
## Year2001          0.3347      0.1602      2.09  0.03687 *
## Year2002          0.2773      0.1483      1.87  0.06164 .
## Year2003          0.3236      0.1629      1.99  0.04716 *
## Year2004          0.2896      0.1471      1.97  0.04925 *
## Year2005          0.1763      0.1541      1.14  0.25267
## Year2006          0.0880      0.1384      0.64  0.52535
## Year2007          0.1083      0.1375      0.79  0.43099
## Year2008          0.0636      0.1334      0.48  0.63343
## Year2009          0.2920      0.1386      2.11  0.03532 *
## Year2010          0.1760      0.1407      1.25  0.21097
## Year2011          0.3966      0.1360      2.92  0.00360 **
## Year2012          0.3063      0.1408      2.18  0.02975 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.708
## Multiple R-squared:  0.0546, Adjusted R-squared:  0.0376
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1132 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.384  0.872  0.948  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          7.99e-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 1.993 1          1.412
## LastAuthorFemale 1.937 1          1.392
## Year          1.187 16          1.005

```

Residuals from first and last author



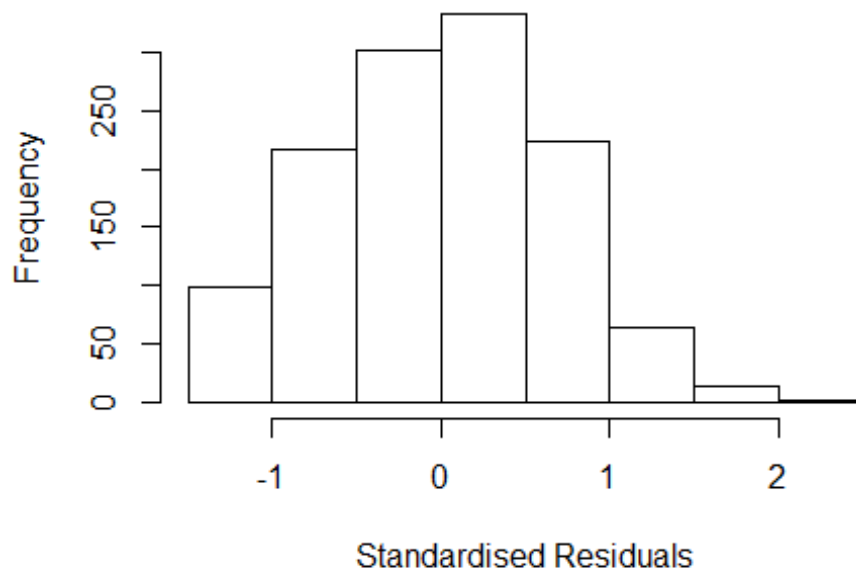
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3371 -0.5062 0.0117 0.4706 2.0609
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8733 0.1168 7.48 1.4e-13 ***
## FirstAuthorFemale1 -0.0329 0.0562 -0.59 0.55853
## LastAuthorFemale1 -0.0195 0.0567 -0.34 0.73080
## Year1997 0.1866 0.1505 1.24 0.21533
## Year1998 0.0705 0.1589 0.44 0.65724
## Year1999 0.1275 0.1333 0.96 0.33876
## Year2000 0.2188 0.1514 1.45 0.14853
## Year2001 0.3811 0.1601 2.38 0.01746 *
## Year2002 0.3246 0.1457 2.23 0.02601 *
## Year2003 0.3670 0.1621 2.26 0.02373 *
## Year2004 0.3212 0.1446 2.22 0.02653 *
## Year2005 0.2255 0.1504 1.50 0.13413
```

```

## Year2006          0.1233      0.1366      0.90  0.36697
## Year2007          0.1666      0.1353      1.23  0.21864
## Year2008          0.1069      0.1303      0.82  0.41235
## Year2009          0.3383      0.1363      2.48  0.01318 *
## Year2010          0.2324      0.1390      1.67  0.09483 .
## Year2011          0.4638      0.1324      3.50  0.00048 ***
## Year2012          0.3635      0.1385      2.62  0.00880 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.0332, Adjusted R-squared:  0.0191
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1138 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.380  0.868  0.947  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      7.99e-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 1.106 1      1.052
## Year              1.106 16      1.003

```

Residuals from first author



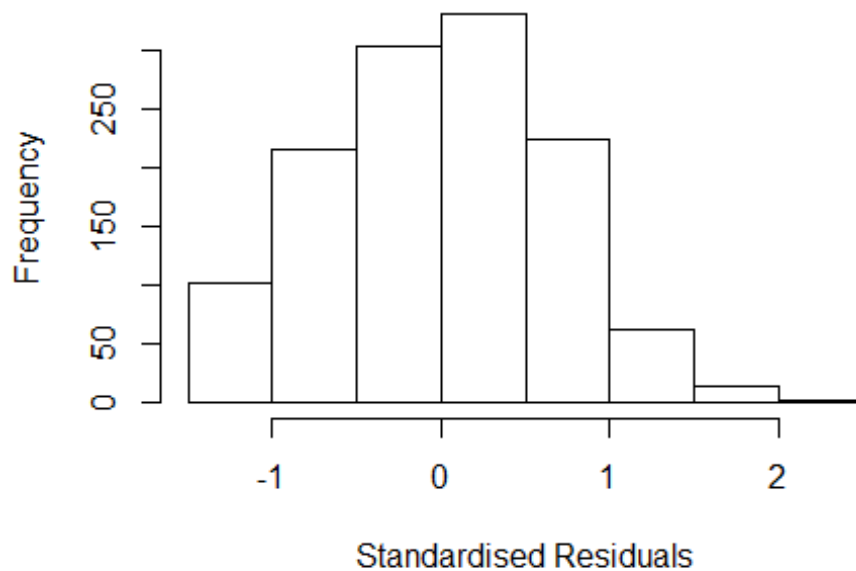
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3342 -0.5073 0.0103 0.4726 2.0638
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8725 0.1166 7.48 1.4e-13 ***
## FirstAuthorFemale1 -0.0464 0.0419 -1.11 0.2688
## Year1997 0.1840 0.1504 1.22 0.2215
## Year1998 0.0707 0.1589 0.44 0.6567
## Year1999 0.1256 0.1334 0.94 0.3466
## Year2000 0.2191 0.1513 1.45 0.1478
## Year2001 0.3811 0.1598 2.38 0.0173 *
## Year2002 0.3227 0.1453 2.22 0.0265 *
## Year2003 0.3646 0.1615 2.26 0.0241 *
## Year2004 0.3185 0.1446 2.20 0.0278 *
## Year2005 0.2234 0.1501 1.49 0.1369
## Year2006 0.1210 0.1366 0.89 0.3756
```

```

## Year2007          0.1655      0.1354      1.22      0.2218
## Year2008          0.1065      0.1303      0.82      0.4141
## Year2009          0.3377      0.1363      2.48      0.0134 *
## Year2010          0.2315      0.1391      1.66      0.0963 .
## Year2011          0.4617      0.1322      3.49      0.0005 ***
## Year2012          0.3629      0.1386      2.62      0.0090 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.0332, Adjusted R-squared:  0.0198
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1135 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.378  0.868  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      7.99e-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 1.076 1          1.038
## Year              1.076 16          1.002

```

Residuals from last author



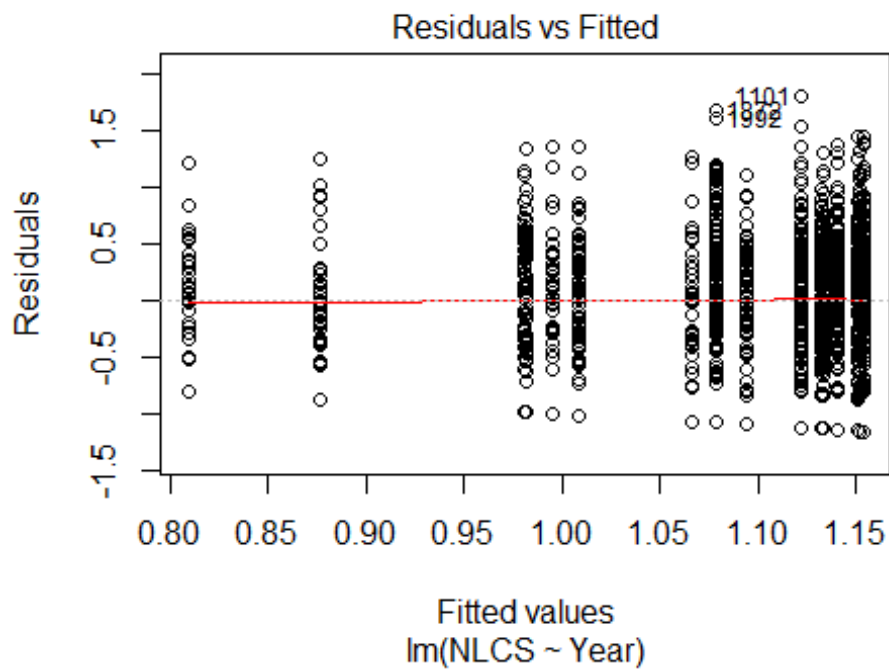
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3325 -0.5066 0.0142 0.4728 2.0655
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8705 0.1172 7.43 2.1e-13 ***
## LastAuthorFemale1 -0.0425 0.0423 -1.00 0.31541
## Year1997 0.1903 0.1506 1.26 0.20652
## Year1998 0.0703 0.1589 0.44 0.65827
## Year1999 0.1295 0.1336 0.97 0.33263
## Year2000 0.2183 0.1515 1.44 0.15000
## Year2001 0.3795 0.1606 2.36 0.01826 *
## Year2002 0.3243 0.1461 2.22 0.02661 *
## Year2003 0.3657 0.1626 2.25 0.02467 *
## Year2004 0.3213 0.1448 2.22 0.02670 *
## Year2005 0.2248 0.1509 1.49 0.13648
## Year2006 0.1232 0.1372 0.90 0.36925
```

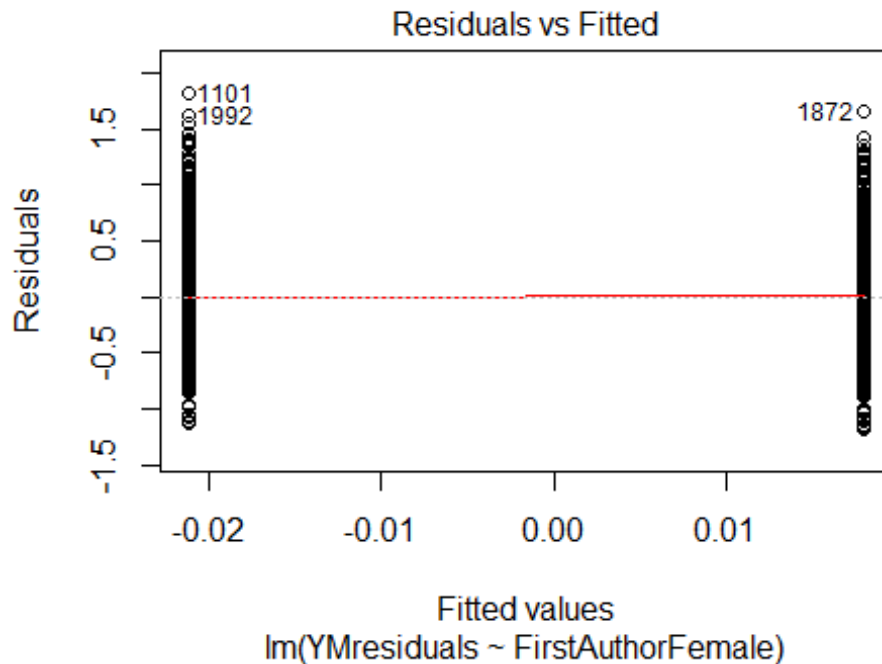
```

## Year2007          0.1627      0.1355      1.20  0.22990
## Year2008          0.1029      0.1303      0.79  0.43006
## Year2009          0.3342      0.1362      2.45  0.01430 *
## Year2010          0.2289      0.1391      1.64  0.10025
## Year2011          0.4620      0.1328      3.48  0.00052 ***
## Year2012          0.3601      0.1384      2.60  0.00938 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.033, Adjusted R-squared:  0.0196
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1137 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.377  0.868  0.947  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      7.99e-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: 1251"
## [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
##   72   56   47   50   74   67   79   48   53   85   86  123  168  189  211
## 2011 2012
##  213  257
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   61   47   36   43   59   51   60   37   43   73   73  106  148  163  176
## 2011 2012

```

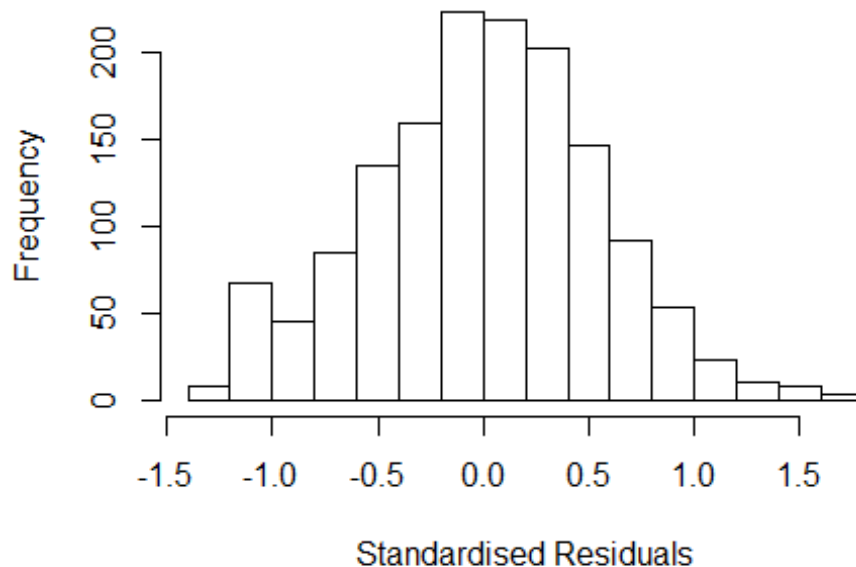
```
## 176 211
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 59 44 33 42 57 48 57 37 41 67 64 93 143 153 169
## 2011 2012
## 171 204
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```





```
## [1] "Female first author team size 2018 geometric mean: 2.21770784496183"
## [1] "Male first author team size 2018 geometric mean: 1.80321561931492"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8100, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.11567096083027"
## [1] "Male last author team size 2018 geometric mean: 1.98791233516702"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7500, 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 1.809 1          1.345
## LastAuthorFemale  1.758 1          1.326
## UniqueAuthors    1.211 4          1.024
## Year             1.301 16          1.008
```

Residuals from first and last author and team size



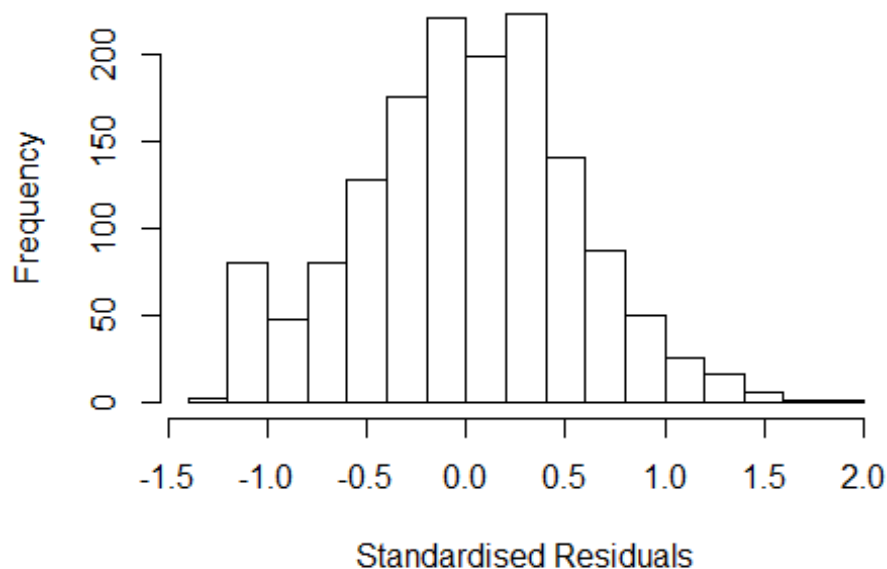
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2883 -0.3659  0.0166  0.3622  1.7174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8363    0.0687   12.17 < 2e-16 ***
## FirstAuthorFemale1 -0.0370    0.0393   -0.94  0.34633
## LastAuthorFemale1  0.0926    0.0385    2.41  0.01625 *
## UniqueAuthors2     0.1135    0.0356    3.19  0.00145 **
## UniqueAuthors3     0.1714    0.0471    3.64  0.00028 ***
## UniqueAuthors4     0.0878    0.0799    1.10  0.27193
## UniqueAuthors5     0.1559    0.0683    2.28  0.02256 *
## Year1997          0.1071    0.1064    1.01  0.31453
## Year1998         -0.1064    0.1160   -0.92  0.35939
## Year1999          0.0723    0.1183    0.61  0.54158
```

```

## Year2000          0.0559      0.1098      0.51  0.61054
## Year2001          0.2376      0.0975      2.44  0.01490 *
## Year2002          0.2249      0.1122      2.00  0.04523 *
## Year2003          0.1463      0.1162      1.26  0.20793
## Year2004          0.2829      0.1123      2.52  0.01189 *
## Year2005          0.1166      0.0920      1.27  0.20504
## Year2006          0.2141      0.0940      2.28  0.02291 *
## Year2007          0.1633      0.0891      1.83  0.06709 .
## Year2008          0.1940      0.0859      2.26  0.02404 *
## Year2009          0.2016      0.0814      2.48  0.01334 *
## Year2010          0.2150      0.0798      2.69  0.00713 **
## Year2011          0.2233      0.0819      2.72  0.00651 **
## Year2012          0.1300      0.0825      1.58  0.11534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0417, Adjusted R-squared:  0.0273
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.281  0.868  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          6.75e-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 1.815 1          1.347
## LastAuthorFemale 1.793 1          1.339
## Year              1.128 16          1.004

```

Residuals from first and last author



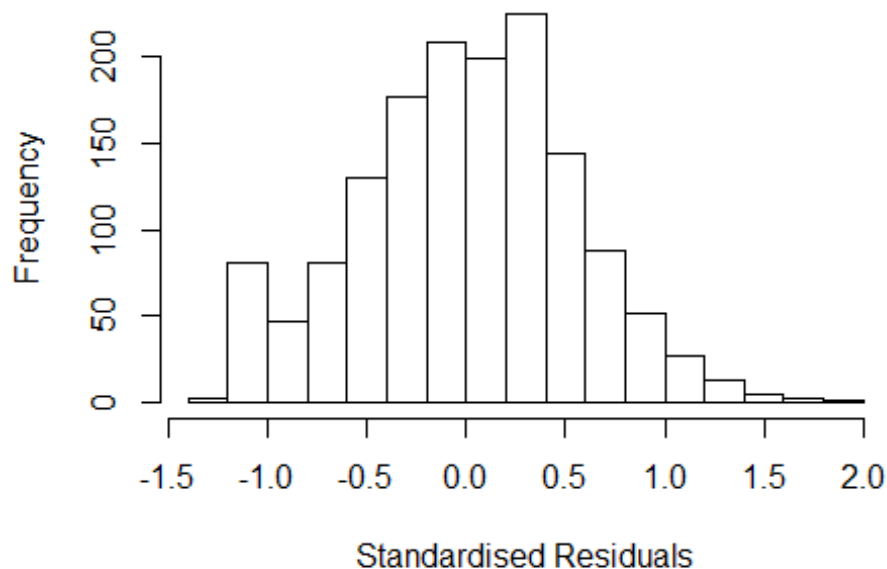
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22747 -0.35002 0.00925 0.35487 1.84221
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8539 0.0692 12.34 <2e-16 ***
## FirstAuthorFemale1 -0.0222 0.0397 -0.56 0.5766
## LastAuthorFemale1 0.0913 0.0392 2.33 0.0201 *
## Year1997 0.1241 0.1048 1.18 0.2365
## Year1998 -0.1121 0.1139 -0.98 0.3250
## Year1999 0.1107 0.1179 0.94 0.3479
## Year2000 0.0738 0.1112 0.66 0.5069
## Year2001 0.2443 0.0980 2.49 0.0127 *
## Year2002 0.2655 0.1108 2.40 0.0167 *
## Year2003 0.1615 0.1162 1.39 0.1650
## Year2004 0.3045 0.1132 2.69 0.0073 **
## Year2005 0.1310 0.0916 1.43 0.1531
```

```

## Year2006          0.2409      0.0952      2.53      0.0115 *
## Year2007          0.2013      0.0891      2.26      0.0239 *
## Year2008          0.2229      0.0858      2.60      0.0095 **
## Year2009          0.2319      0.0813      2.85      0.0044 **
## Year2010          0.2546      0.0801      3.18      0.0015 **
## Year2011          0.2641      0.0824      3.21      0.0014 **
## Year2012          0.1690      0.0831      2.03      0.0422 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.0285, Adjusted R-squared:  0.0165
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1345 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.216  0.866  0.951  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      6.75e-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 1.083 1      1.041
## Year              1.083 16      1.002

```

Residuals from first author



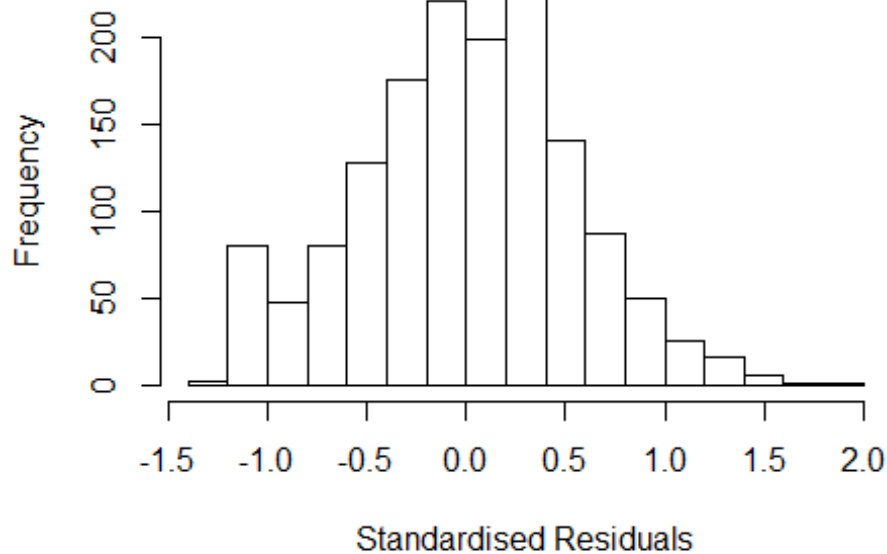
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20717 -0.35570 0.00712 0.35190 1.82988
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8629 0.0697 12.39 <2e-16 ***
## FirstAuthorFemale1 0.0381 0.0308 1.24 0.2165
## Year1997 0.1352 0.1056 1.28 0.2003
## Year1998 -0.1098 0.1139 -0.96 0.3355
## Year1999 0.1126 0.1188 0.95 0.3431
## Year2000 0.0804 0.1119 0.72 0.4725
## Year2001 0.2519 0.0989 2.55 0.0110 *
## Year2002 0.2676 0.1116 2.40 0.0166 *
## Year2003 0.1684 0.1186 1.42 0.1558
## Year2004 0.3062 0.1143 2.68 0.0075 **
## Year2005 0.1336 0.0922 1.45 0.1477
## Year2006 0.2563 0.0947 2.71 0.0069 **
```

```

## Year2007          0.2102      0.0886      2.37      0.0179 *
## Year2008          0.2262      0.0863      2.62      0.0088 **
## Year2009          0.2349      0.0817      2.87      0.0041 **
## Year2010          0.2608      0.0807      3.23      0.0013 **
## Year2011          0.2698      0.0832      3.24      0.0012 **
## Year2012          0.1799      0.0835      2.16      0.0312 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0248, Adjusted R-squared:  0.0135
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 142 weights are ~= 1. The remaining 1340 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.221  0.869  0.950  0.903  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      6.75e-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 1.068 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
## -1.23127 -0.35272 0.00866 0.35360 1.84758
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8516 0.0692 12.30 <2e-16 ***
## LastAuthorFemale1 0.0770 0.0303 2.54 0.0111 *
## Year1997 0.1230 0.1047 1.17 0.2405
## Year1998 -0.1126 0.1141 -0.99 0.3241
## Year1999 0.1087 0.1180 0.92 0.3571
## Year2000 0.0743 0.1112 0.67 0.5039
## Year2001 0.2424 0.0979 2.48 0.0134 *
## Year2002 0.2624 0.1111 2.36 0.0184 *
## Year2003 0.1584 0.1163 1.36 0.1734
## Year2004 0.3027 0.1131 2.68 0.0075 **
## Year2005 0.1288 0.0915 1.41 0.1598
## Year2006 0.2414 0.0952 2.53 0.0114 *
```

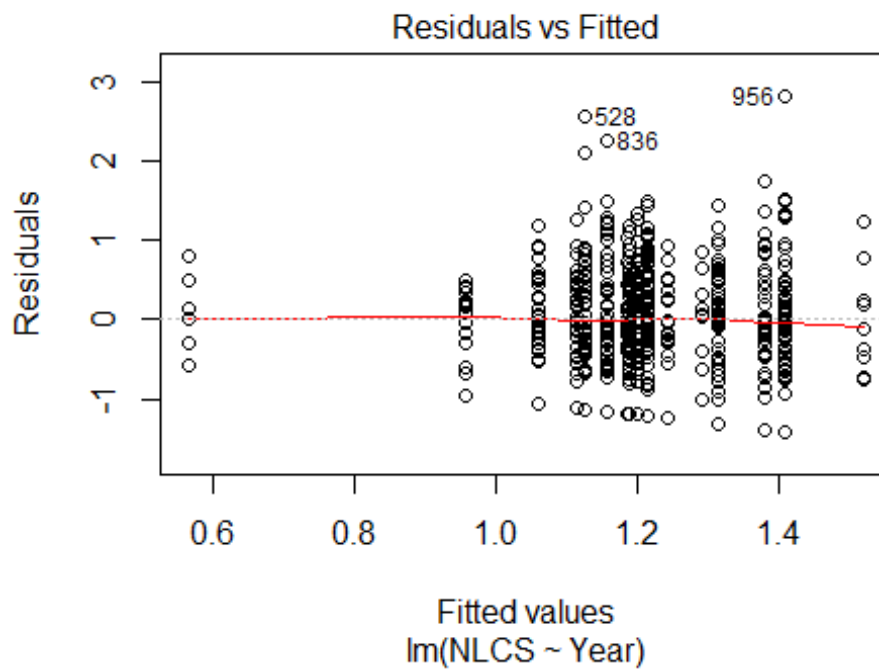


```

## Year2007          0.1990      0.0888      2.24      0.0251 *
## Year2008          0.2198      0.0856      2.57      0.0103 *
## Year2009          0.2278      0.0809      2.81      0.0049 **
## Year2010          0.2516      0.0802      3.14      0.0017 **
## Year2011          0.2612      0.0821      3.18      0.0015 **
## Year2012          0.1664      0.0831      2.00      0.0454 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0283, Adjusted R-squared:  0.017
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.209  0.867  0.951  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      6.75e-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: 1482"
## [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
##   19   24   27   16   13   25   51   62   45   57   64   57   50   52   47
## 2011 2012
##   76   73
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9   15   18    9    7   16   40   50   39   50   57   53   36   44   39
## 2011 2012

```

```
## 68 58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 13 17 9 7 15 38 48 38 47 54 51 36 42 38
## 2011 2012
## 65 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 = 25, df = 16, p-value = 0.08
```



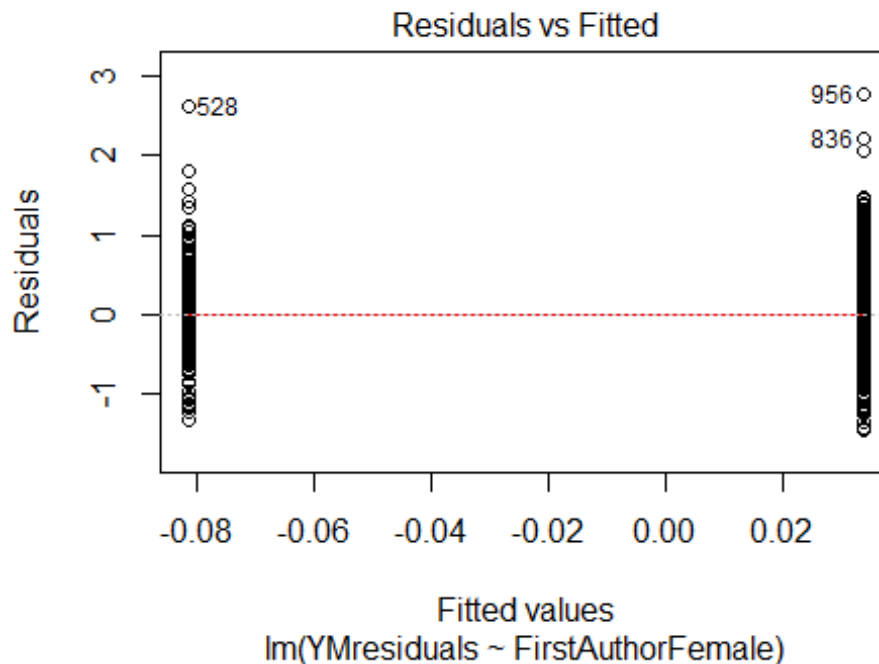
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.22, df = 1, p-value = 0.6

## [1] "Female first author team size 2018 geometric mean: 2.34930416577067"
## [1] "Male first author team size 2018 geometric mean: 2.10254774888131"

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

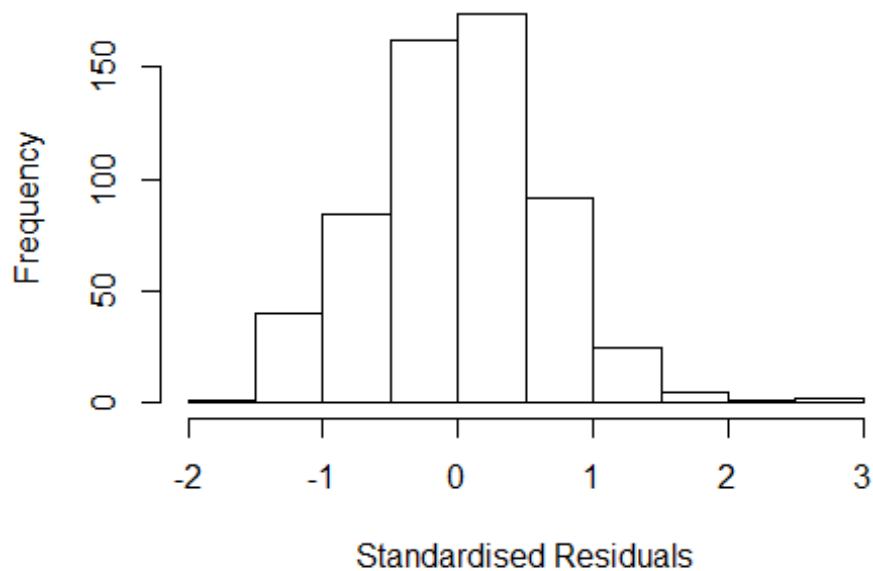
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 610, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.21368777146356"
## [1] "Male last author team size 2018 geometric mean: 2.2720044690055"

## 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 = 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^(1/(2*Df))
## FirstAuthorFemale 1.797  1      1.340
## LastAuthorFemale  2.008  1      1.417
## UniqueAuthors    1.858  4      1.081
## Year              2.067 16      1.023
```

Residuals from first and last author and team size



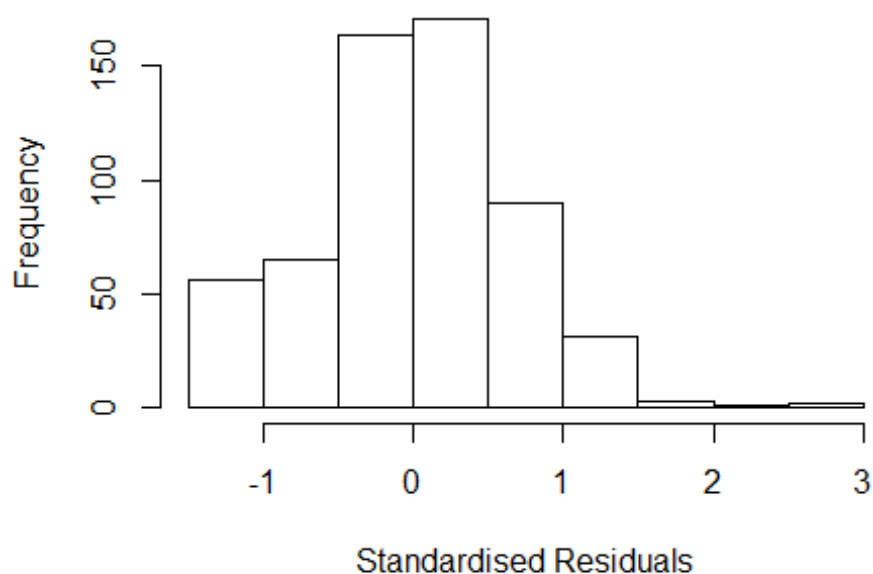
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 528 34848880511 3.674 2007    1200     2    2.861
## 956 83655184689 4.215 2012    1202     3    2.945
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5004 -0.4234  0.0132  0.4274  2.9446
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.3396    0.2381    5.63 2.9e-08 ***
## FirstAuthorFemale1 -0.1005    0.0757   -1.33  0.1851
## LastAuthorFemale1 -0.0583    0.0803   -0.73  0.4683
## UniqueAuthors2     0.2299    0.0748    3.07  0.0022 **
## UniqueAuthors3     0.4167    0.0902    4.62 4.8e-06 ***
## UniqueAuthors4     0.4021    0.1220    3.30  0.0010 **
## UniqueAuthors5     0.2525    0.1208    2.09  0.0370 *
## Year1997        -0.2023    0.2698   -0.75  0.4536
## Year1998        -0.1754    0.2639   -0.66  0.5065
```

```

## Year1999          -0.1746      0.3435    -0.51    0.6115
## Year2000          -0.8030      0.2978    -2.70    0.0072 **
## Year2001          -0.3971      0.2596    -1.53    0.1266
## Year2002           0.0334      0.2719     0.12    0.9024
## Year2003          -0.2096      0.2565    -0.82    0.4142
## Year2004          -0.2380      0.2581    -0.92    0.3568
## Year2005          -0.1632      0.2660    -0.61    0.5398
## Year2006          -0.0778      0.2547    -0.31    0.7600
## Year2007          -0.3673      0.2582    -1.42    0.1554
## Year2008          -0.1169      0.2598    -0.45    0.6529
## Year2009          -0.2491      0.2575    -0.97    0.3339
## Year2010          -0.3292      0.2616    -1.26    0.2088
## Year2011          -0.3471      0.2565    -1.35    0.1766
## Year2012          -0.0692      0.2579    -0.27    0.7886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.645
## Multiple R-squared:  0.0964, Adjusted R-squared:  0.0609
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 533 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0026 0.8660 0.9520 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          1.72e-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 1.826 1          1.351
## LastAuthorFemale 1.929 1          1.389
## Year          1.329 16          1.009

```

Residuals from first and last author



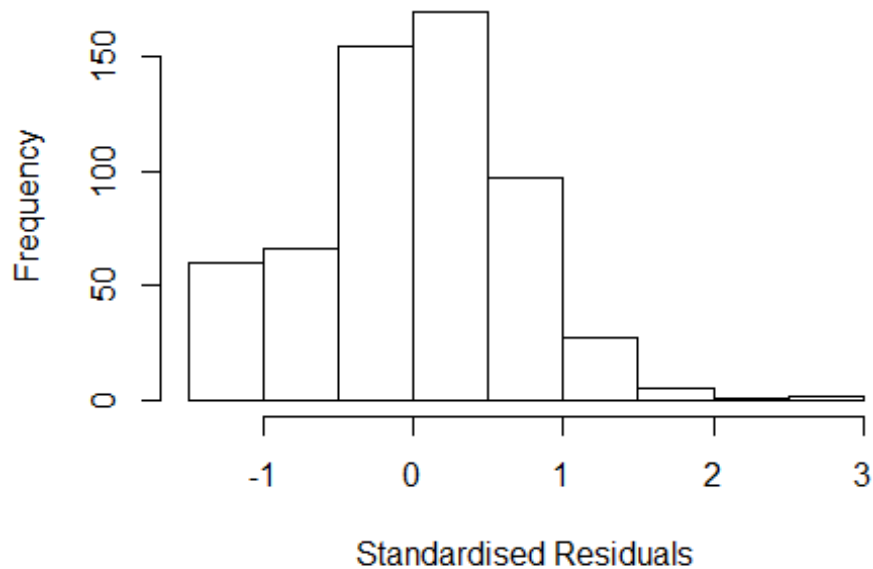
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 528 34848880511 3.674 2007    1200    2    2.755
## 956 83655184689 4.215 2012    1202    3    2.813
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4503 -0.4128  0.0124  0.4436  2.8131
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4799     0.2316   6.39 3.5e-10 ***
## FirstAuthorFemale1 -0.0570     0.0794  -0.72  0.4728
## LastAuthorFemale1 -0.0812     0.0836  -0.97  0.3316
## Year1997          -0.2243     0.2705  -0.83  0.4074
## Year1998          -0.1907     0.2619  -0.73  0.4669
## Year1999          -0.1637     0.3064  -0.53  0.5934
## Year2000          -0.8784     0.3044  -2.89  0.0041 **
## Year2001          -0.4823     0.2560  -1.88  0.0601 .
## Year2002          -0.0296     0.2666  -0.11  0.9117
## Year2003          -0.2858     0.2479  -1.15  0.2494
## Year2004          -0.3222     0.2540  -1.27  0.2051
```

```

## Year2005          -0.2029      0.2686   -0.76   0.4504
## Year2006          -0.1355      0.2506   -0.54   0.5889
## Year2007          -0.4222      0.2588   -1.63   0.1034
## Year2008          -0.1968      0.2550   -0.77   0.4406
## Year2009          -0.2156      0.2501   -0.86   0.3890
## Year2010          -0.3659      0.2547   -1.44   0.1514
## Year2011          -0.3516      0.2500   -1.41   0.1601
## Year2012          -0.0780      0.2525   -0.31   0.7574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0499, Adjusted R-squared:  0.0196
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 532 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0216 0.8700 0.9520 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      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 1.158 1      1.076
## Year      1.158 16      1.005

```

Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 528 34848880511 3.674 2007    1200    2    2.755
## 956 83655184689 4.215 2012    1202    3    2.813
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4463 -0.4169  0.0216  0.4583  2.8323
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4799     0.2316   6.39 3.5e-10 ***
## FirstAuthorFemale1 -0.1046     0.0636  -1.65  0.1004
## Year1997         -0.2267     0.2695  -0.84  0.4007
## Year1998         -0.1948     0.2619  -0.74  0.4573
## Year1999         -0.1748     0.3053  -0.57  0.5672
## Year2000         -0.8907     0.3051  -2.92  0.0036 **
## Year2001         -0.4965     0.2558  -1.94  0.0528 .
## Year2002         -0.0336     0.2664  -0.13  0.8995
## Year2003         -0.2884     0.2476  -1.16  0.2446
## Year2004         -0.3339     0.2534  -1.32  0.1882
## Year2005         -0.2190     0.2671  -0.82  0.4126
```

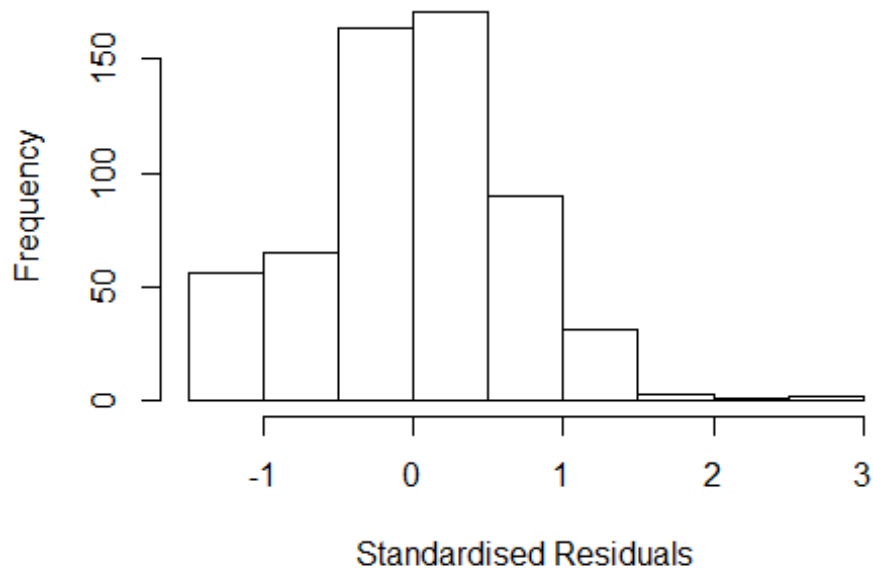


```

## Year2006          -0.1429      0.2504   -0.57   0.5683
## Year2007          -0.4195      0.2591   -1.62   0.1060
## Year2008          -0.2060      0.2550   -0.81   0.4195
## Year2009          -0.2157      0.2505   -0.86   0.3895
## Year2010          -0.3669      0.2547   -1.44   0.1502
## Year2011          -0.3705      0.2495   -1.49   0.1380
## Year2012          -0.0972      0.2518   -0.39   0.6997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0482, Adjusted R-squared:  0.0196
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 527 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0184 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      1.72e-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 1.222 1      1.105
## Year              1.222 16      1.006

```

Residuals from last author



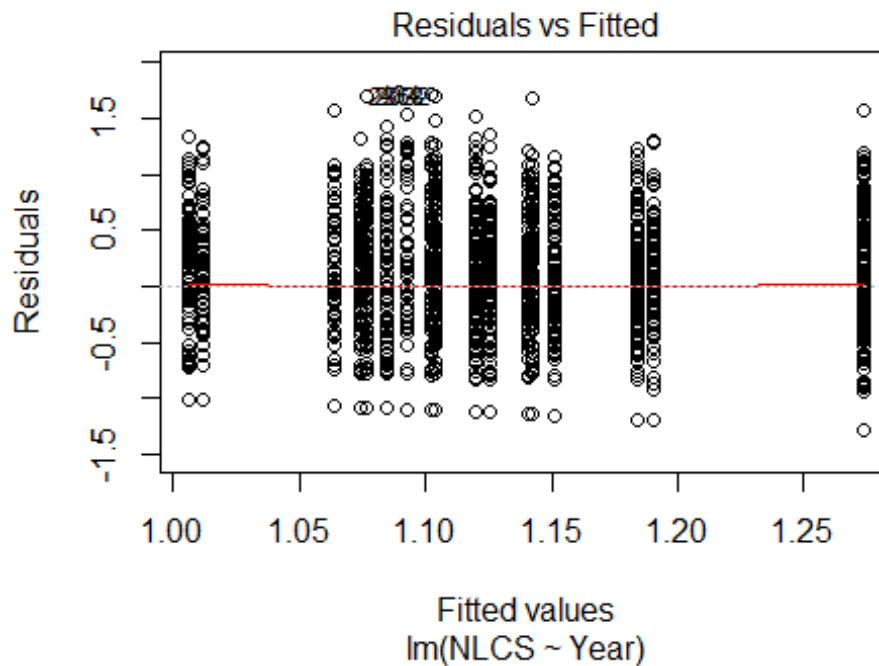
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 528 34848880511 3.674 2007    1200    2    2.755
## 956 83655184689 4.215 2012    1202    3    2.813
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4370 -0.4105  0.0168  0.4488  2.8214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4797     0.2317   6.39 3.5e-10 ***
## LastAuthorFemale1 -0.1174     0.0667  -1.76  0.0791 .
## Year1997         -0.2296     0.2704  -0.85  0.3963
## Year1998         -0.1988     0.2623  -0.76  0.4489
## Year1999         -0.1586     0.3066  -0.52  0.6051
## Year2000         -0.8839     0.3029  -2.92  0.0037 **
## Year2001         -0.4845     0.2564  -1.89  0.0593 .
## Year2002         -0.0427     0.2654  -0.16  0.8722
## Year2003         -0.2928     0.2480  -1.18  0.2382
## Year2004         -0.3269     0.2540  -1.29  0.1986
## Year2005         -0.2110     0.2685  -0.79  0.4323
```

```

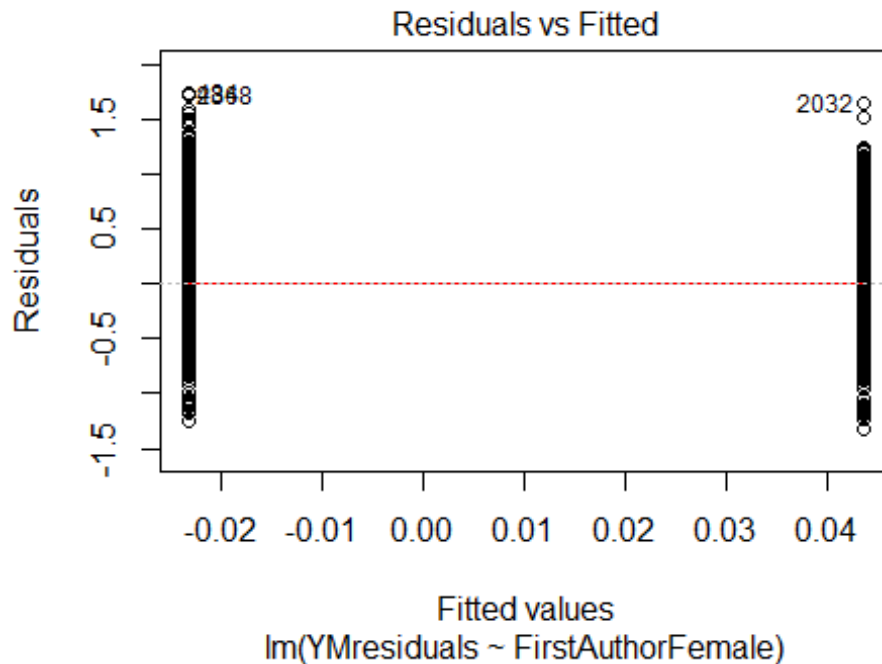
## Year2006          -0.1389      0.2506   -0.55   0.5795
## Year2007          -0.4299      0.2583   -1.66   0.0966 .
## Year2008          -0.2022      0.2552   -0.79   0.4284
## Year2009          -0.2302      0.2500   -0.92   0.3575
## Year2010          -0.3761      0.2544   -1.48   0.1399
## Year2011          -0.3551      0.2499   -1.42   0.1558
## Year2012          -0.0861      0.2518   -0.34   0.7324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0203
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 533 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0195 0.8680 0.9510 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      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 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
##   95   99  123  116  115  159  177  139  129  146  187  170  179  217  228
## 2011 2012
##  165  261
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   69   76   92   93   85   64  135  109  116  124  158  141  147  173  190

```

```
## 2011 2012
## 130 202
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 68 70 88 86 83 60 130 104 113 120 148 132 137 167 182
## 2011 2012
## 126 190
## [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
```

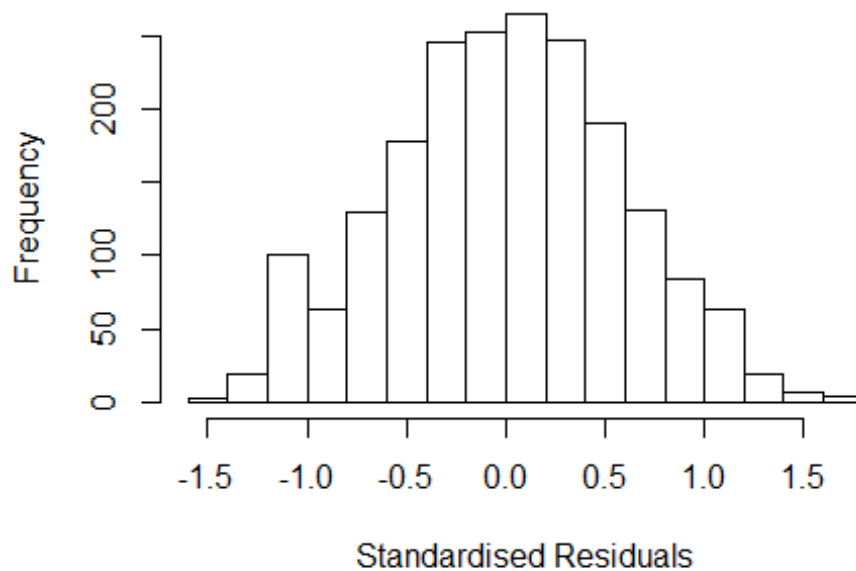


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.9, df = 1, p-value = 0.09
```



```
## [1] "Female first author team size 2018 geometric mean: 1.94141792047439"
## [1] "Male first author team size 2018 geometric mean: 1.56637985729753"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.73263141829486"
## [1] "Male last author team size 2018 geometric mean: 1.84045372543658"
##
## 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 2.037 1      1.427
## LastAuthorFemale  2.014 1      1.419
## UniqueAuthors    1.273 4      1.031
## Year             1.283 16      1.008
```

Residuals from first and last author and team size



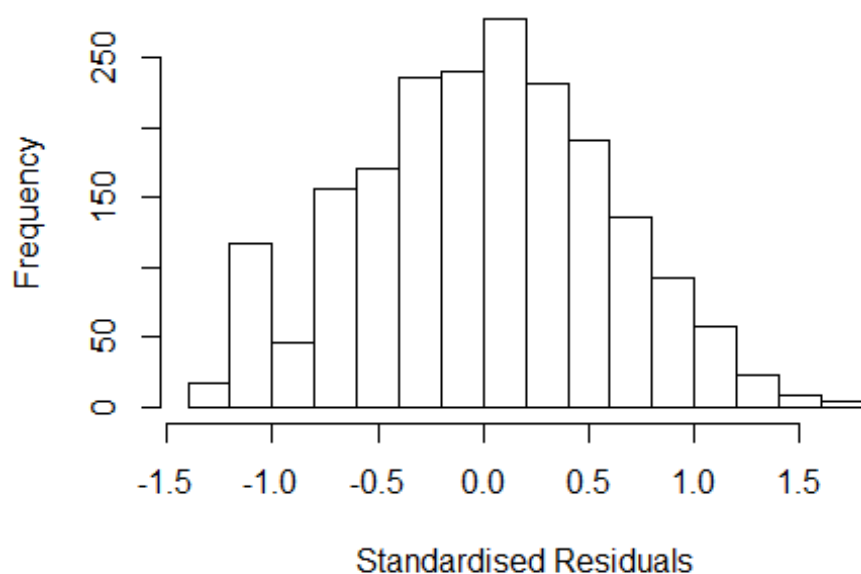
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.51072 -0.39598  0.00564  0.39803  1.75804
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.01280    0.08323   12.17 < 2e-16 ***
## FirstAuthorFemale1  0.08253    0.03968    2.08  0.038 *
## LastAuthorFemale1 -0.02585    0.04035   -0.64  0.522
## UniqueAuthors2     0.20986    0.03619    5.80 7.7e-09 ***
## UniqueAuthors3     0.22930    0.04668    4.91 9.7e-07 ***
## UniqueAuthors4     0.06636    0.10677    0.62  0.534
## UniqueAuthors5    -0.03015    0.09098   -0.33  0.740
## Year1997          -0.04729    0.10989   -0.43  0.667
## Year1998           0.00656    0.11245    0.06  0.954
## Year1999           0.09593    0.10489    0.91  0.361
```

```

## Year2000      0.00481    0.11552    0.04    0.967
## Year2001     -0.10312    0.14402   -0.72    0.474
## Year2002     -0.08169    0.09972   -0.82    0.413
## Year2003      0.05274    0.09676    0.55    0.586
## Year2004      0.06640    0.09703    0.68    0.494
## Year2005     -0.02294    0.09777   -0.23    0.814
## Year2006      0.03084    0.09585    0.32    0.748
## Year2007     -0.00499    0.09457   -0.05    0.958
## Year2008     -0.06153    0.09865   -0.62    0.533
## Year2009      0.07945    0.09208    0.86    0.388
## Year2010     -0.01507    0.09460   -0.16    0.873
## Year2011      0.03678    0.10159    0.36    0.717
## Year2012      0.18609    0.09611    1.94    0.053 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.597
## Multiple R-squared:  0.0443, Adjusted R-squared:  0.0337
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 155 weights are ~= 1. The remaining 1849 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.365  0.875  0.951  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      4.99e-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 2.211 1      1.487
## LastAuthorFemale 2.171 1      1.473
## 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
## -1.3652 -0.4034 0.0123 0.4100 1.7439
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03917 0.08367 12.42 <2e-16 ***
## FirstAuthorFemale1 0.10794 0.04225 2.55 0.011 *
## LastAuthorFemale1 -0.04260 0.04291 -0.99 0.321
## Year1997 -0.04249 0.10816 -0.39 0.694
## Year1998 0.01330 0.11387 0.12 0.907
## Year1999 0.11744 0.10544 1.11 0.265
## Year2000 0.02894 0.11802 0.25 0.806
## Year2001 -0.04735 0.14697 -0.32 0.747
## Year2002 -0.06207 0.10097 -0.61 0.539
## Year2003 0.08911 0.09748 0.91 0.361
## Year2004 0.09346 0.09893 0.94 0.345
## Year2005 0.01235 0.09787 0.13 0.900
```

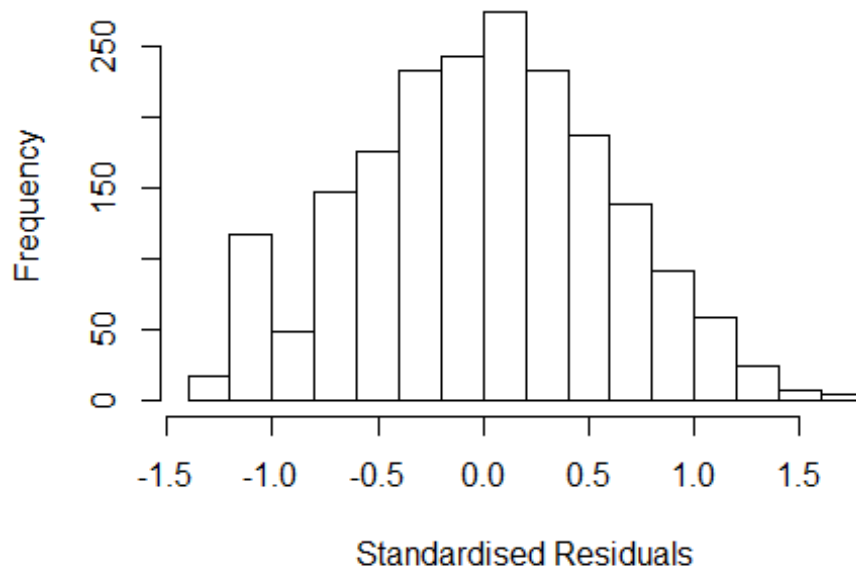


```

## Year2006          0.06205      0.09648      0.64      0.520
## Year2007          0.03338      0.09613      0.35      0.728
## Year2008         -0.00945      0.09884     -0.10      0.924
## Year2009          0.11436      0.09258      1.24      0.217
## Year2010          0.01995      0.09575      0.21      0.835
## Year2011          0.08027      0.10171      0.79      0.430
## Year2012          0.21804      0.09676      2.25      0.024 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 164 weights are ~= 1. The remaining 1840 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.389  0.871  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      4.99e-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 1.055 1      1.027
## Year              1.055 16      1.002

```

Residuals from first author



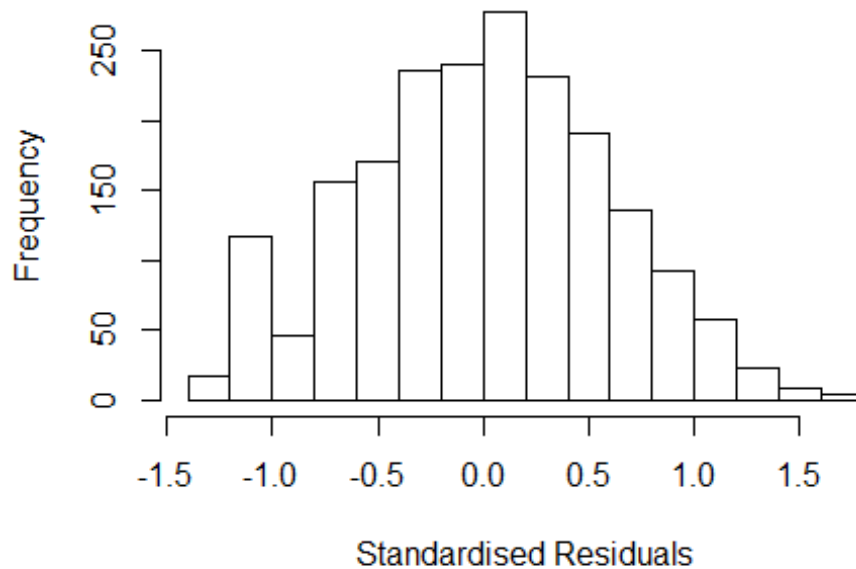
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3305 -0.4065  0.0144  0.4114  1.7448
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0353     0.0836   12.39  <2e-16 ***
## FirstAuthorFemale1  0.0776     0.0293    2.65   0.0081 **
## Year1997         -0.0423     0.1084   -0.39   0.6961
## Year1998          0.0128     0.1138    0.11   0.9105
## Year1999          0.1193     0.1057    1.13   0.2590
## Year2000          0.0319     0.1182    0.27   0.7874
## Year2001         -0.0459     0.1472   -0.31   0.7554
## Year2002         -0.0603     0.1010   -0.60   0.5502
## Year2003          0.0896     0.0975    0.92   0.3585
## Year2004          0.0952     0.0990    0.96   0.3364
## Year2005          0.0129     0.0978    0.13   0.8952
## Year2006          0.0635     0.0965    0.66   0.5103
```

```

## Year2007          0.0331      0.0961      0.34      0.7305
## Year2008         -0.0101      0.0989     -0.10      0.9183
## Year2009          0.1143      0.0926      1.23      0.2173
## Year2010          0.0185      0.0958      0.19      0.8469
## Year2011          0.0803      0.1018      0.79      0.4300
## Year2012          0.2177      0.0968      2.25      0.0247 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.019, Adjusted R-squared:  0.0107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 166 weights are ~= 1. The remaining 1838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.389  0.872  0.951  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      4.99e-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 1.034 1      1.017
## Year              1.034 16      1.001

```

Residuals from last author



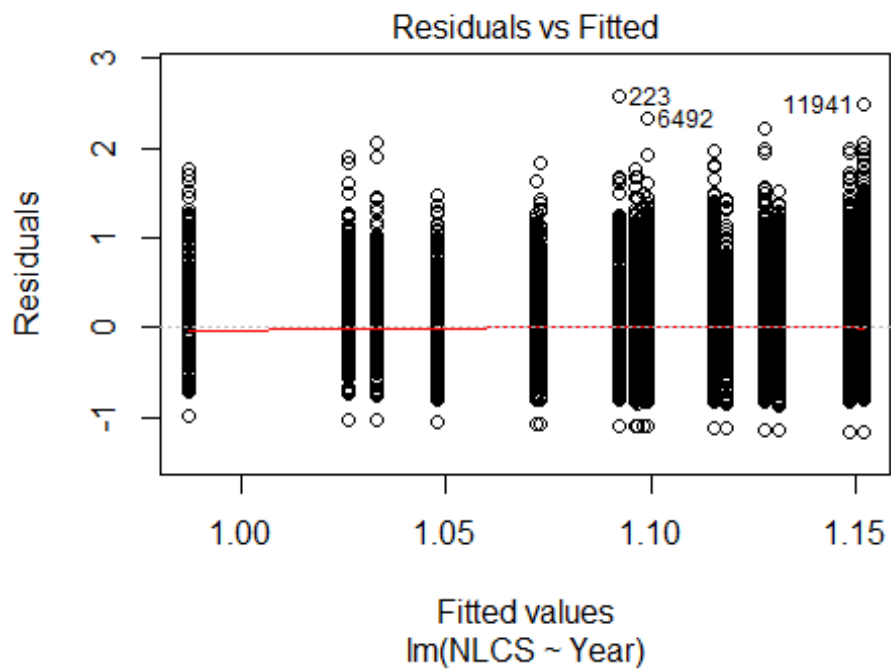
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3101 -0.4068 0.0143 0.4129 1.7364
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04477 0.08335 12.53 <2e-16 ***
## LastAuthorFemale1 0.03622 0.02952 1.23 0.220
## Year1997 -0.04759 0.10839 -0.44 0.661
## Year1998 0.01280 0.11310 0.11 0.910
## Year1999 0.12372 0.10595 1.17 0.243
## Year2000 0.03079 0.11791 0.26 0.794
## Year2001 -0.03766 0.14689 -0.26 0.798
## Year2002 -0.05329 0.10086 -0.53 0.597
## Year2003 0.09426 0.09749 0.97 0.334
## Year2004 0.10253 0.09870 1.04 0.299
## Year2005 0.02057 0.09742 0.21 0.833
## Year2006 0.06756 0.09629 0.70 0.483
```

```

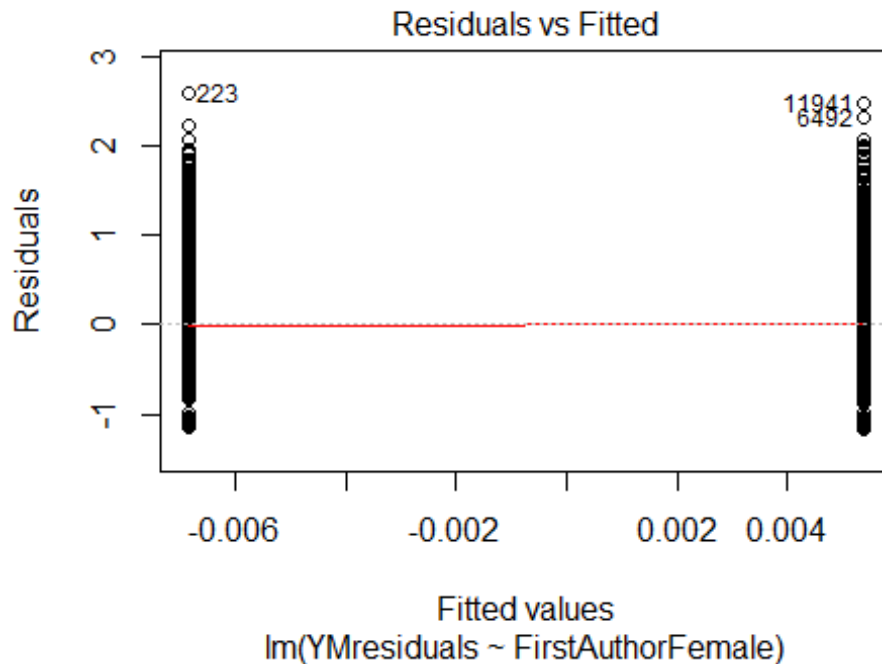
## Year2007          0.04100      0.09575      0.43      0.669
## Year2008         -0.00145      0.09854     -0.01      0.988
## Year2009          0.12158      0.09221      1.32      0.187
## Year2010          0.02375      0.09542      0.25      0.803
## Year2011          0.08835      0.10148      0.87      0.384
## Year2012          0.22907      0.09629      2.38      0.017 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.00791
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1834 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.395  0.871  0.951  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      4.99e-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: 2004"
## [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
## 338 370 410 393 403 462 434 413 476 522 602 731 953 1162 1230
## 2011 2012
## 1356 1540
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 288 301 326 319 327 374 351 351 402 469 510 619 816 993 1050
## 2011 2012

```

```
## 1157 1322
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 277 285 313 311 311 356 334 337 375 447 480 578 760 923 975
## 2011 2012
## 1080 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 = 80, df = 16, p-value = 1e-10
```

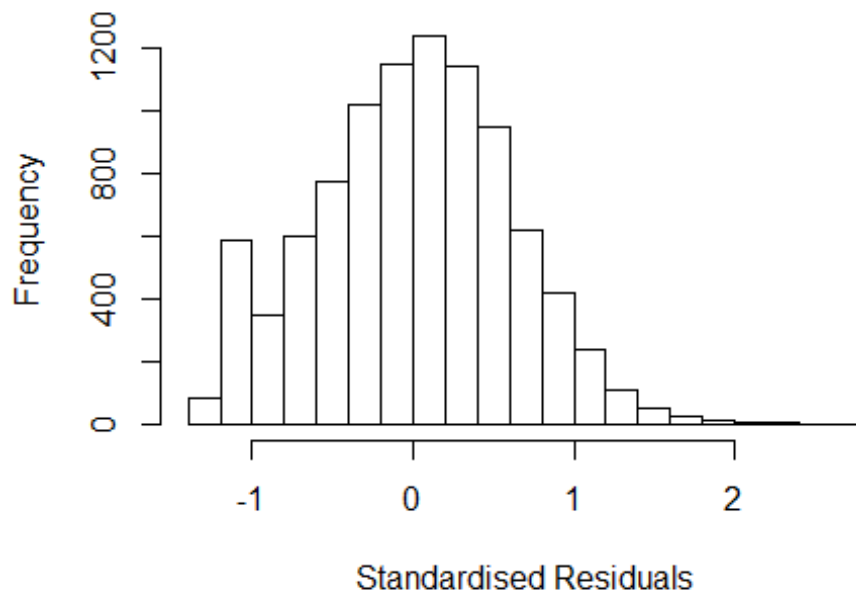


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 24, df = 1, p-value = 1e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 2.13402817001955"
## [1] "Male first author team size 2018 geometric mean: 1.91597531393073"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190000, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.05330801986766"
## [1] "Male last author team size 2018 geometric mean: 2.06391132144101"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190000, 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 1.391 1          1.180
## LastAuthorFemale  1.380 1          1.175
## UniqueAuthors    1.051 4          1.006
## Year              1.081 16         1.002
```

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
## 223 21444433318 3.666 1996      3304      1      2.663
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3136 -0.4075  0.0155  0.4161  2.6630
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.00295    0.04563   21.98 < 2e-16 ***
## FirstAuthorFemale1 0.00462    0.01556    0.30  0.766
## LastAuthorFemale1 0.01283    0.01527    0.84  0.401
## UniqueAuthors2    0.13406    0.01550    8.65 < 2e-16 ***
## UniqueAuthors3    0.20961    0.01897   11.05 < 2e-16 ***
## UniqueAuthors4    0.22954    0.02939    7.81 6.3e-15 ***
## UniqueAuthors5    0.26413    0.03298    8.01 1.3e-15 ***
## Year1997          -0.10416    0.06249   -1.67  0.096 .
## Year1998          -0.07836    0.05790   -1.35  0.176
## Year1999          -0.05504    0.05827   -0.94  0.345
```

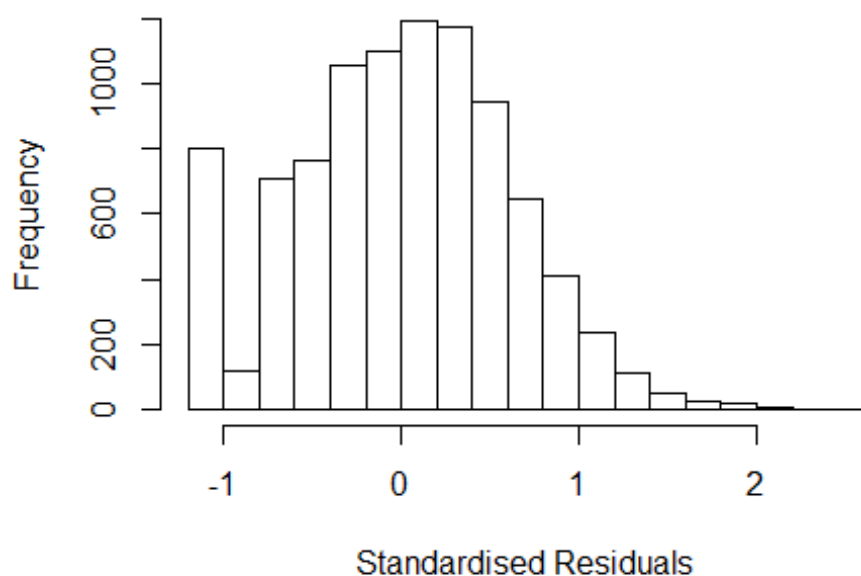


```

## Year2000      0.00550    0.05652    0.10    0.922
## Year2001     -0.01288    0.05566   -0.23    0.817
## Year2002     -0.00892    0.05562   -0.16    0.873
## Year2003     -0.04471    0.05703   -0.78    0.433
## Year2004      0.00724    0.05345    0.14    0.892
## Year2005     -0.02617    0.05301   -0.49    0.621
## Year2006      0.00118    0.05209    0.02    0.982
## Year2007      0.04193    0.05058    0.83    0.407
## Year2008     -0.01607    0.04972   -0.32    0.747
## Year2009     -0.00542    0.04981   -0.11    0.913
## Year2010      0.00827    0.04944    0.17    0.867
## Year2011      0.03002    0.04922    0.61    0.542
## Year2012      0.02051    0.04916    0.42    0.677
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.612
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.0243
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 771 weights are ~= 1. The remaining 8600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0186 0.8720 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          1.07e-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 1.404 1 1.185
## LastAuthorFemale 1.403 1 1.185
## Year 1.045 16 1.001

```

Residuals from first and last author



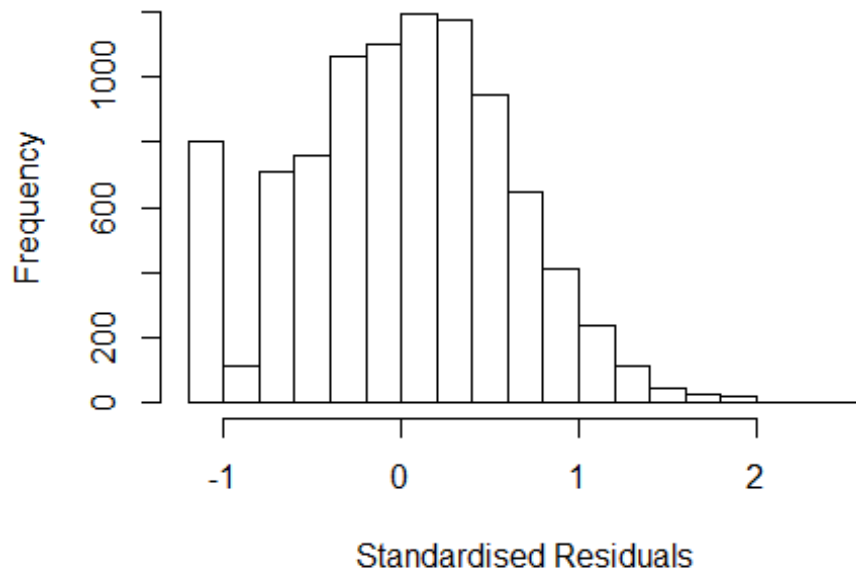
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 223 21444433318 3.666 1996    3304    1    2.6
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1525 -0.4118  0.0215  0.4208  2.5999
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06609    0.04547   23.45  <2e-16 ***
## FirstAuthorFemale1  0.02671    0.01585    1.69   0.092 .
## LastAuthorFemale1  0.00137    0.01560    0.09   0.930
## Year1997        -0.12107    0.06319   -1.92   0.055 .
## Year1998        -0.07129    0.05817   -1.23   0.220
## Year1999        -0.05427    0.05843   -0.93   0.353
## Year2000         0.01600    0.05750    0.28   0.781
## Year2001        -0.00467    0.05624   -0.08   0.934
## Year2002        -0.00923    0.05580   -0.17   0.869
## Year2003        -0.02885    0.05733   -0.50   0.615
## Year2004         0.02612    0.05330    0.49   0.624
## Year2005        -0.01089    0.05311   -0.20   0.838
```

```

## Year2006          0.01886    0.05217    0.36    0.718
## Year2007          0.05828    0.05063    1.15    0.250
## Year2008          0.00358    0.04984    0.07    0.943
## Year2009          0.01611    0.04995    0.32    0.747
## Year2010          0.02881    0.04947    0.58    0.560
## Year2011          0.05463    0.04922    1.11    0.267
## Year2012          0.05527    0.04914    1.12    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.00513,    Adjusted R-squared:  0.00322
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 774 weights are ~= 1. The remaining 8597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0387 0.8720 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      1.07e-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 1.032 1      1.016
## Year              1.032 16      1.001

```

Residuals from first author



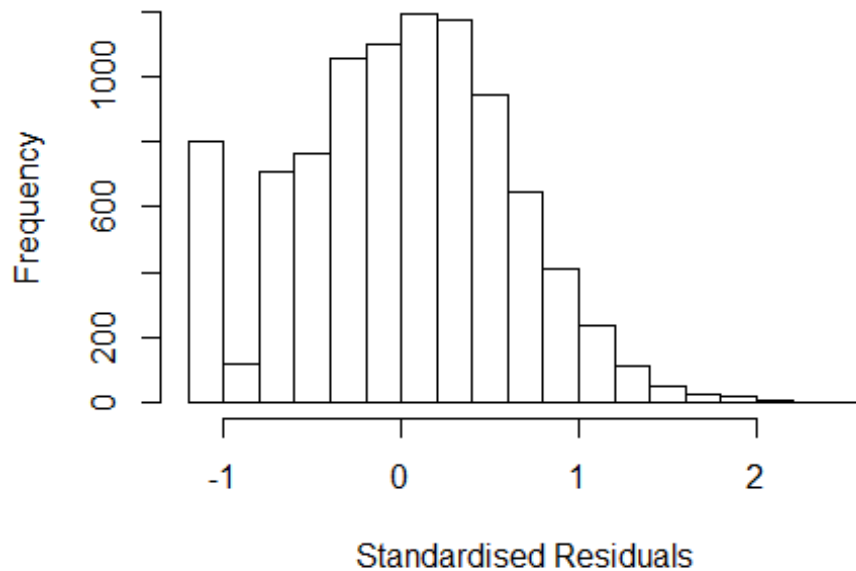
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 223 21444433318 3.666 1996      3304      1      2.6
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1522 -0.4120  0.0218  0.4205  2.5997
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06629    0.04537   23.50  <2e-16 ***
## FirstAuthorFemale1 0.02749    0.01361    2.02   0.043 *
## Year1997      -0.12105    0.06318   -1.92   0.055 .
## Year1998      -0.07127    0.05816   -1.23   0.220
## Year1999      -0.05422    0.05843   -0.93   0.353
## Year2000       0.01608    0.05748    0.28   0.780
## Year2001      -0.00462    0.05624   -0.08   0.935
## Year2002      -0.00917    0.05579   -0.16   0.869
## Year2003      -0.02886    0.05732   -0.50   0.615
## Year2004       0.02617    0.05330    0.49   0.623
## Year2005      -0.01079    0.05310   -0.20   0.839
## Year2006       0.01891    0.05217    0.36   0.717
```

```

## Year2007          0.05838    0.05061    1.15    0.249
## Year2008          0.00370    0.04982    0.07    0.941
## Year2009          0.01622    0.04994    0.32    0.745
## Year2010          0.02895    0.04943    0.59    0.558
## Year2011          0.05475    0.04919    1.11    0.266
## Year2012          0.05538    0.04912    1.13    0.260
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.00513,    Adjusted R-squared:  0.00332
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 774 weights are ~= 1. The remaining 8597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0388 0.8720 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      1.07e-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 1.031 1          1.016
## Year            1.031 16          1.001

```

Residuals from last author



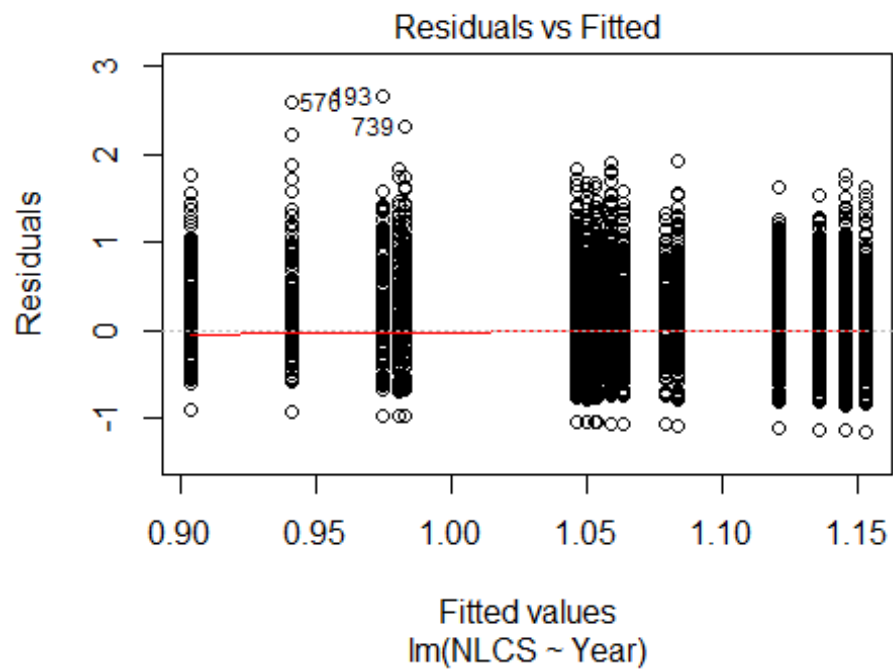
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 223 21444433318 3.666 1996      3304      1      2.6
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1483 -0.4104  0.0249  0.4193  2.5943
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07166    0.04540   23.61  <2e-16 ***
## LastAuthorFemale1 0.01620    0.01340    1.21   0.227
## Year1997       -0.12173    0.06325   -1.92   0.054 .
## Year1998       -0.07161    0.05830   -1.23   0.219
## Year1999       -0.05437    0.05849   -0.93   0.353
## Year2000        0.01669    0.05757    0.29   0.772
## Year2001       -0.00435    0.05632   -0.08   0.938
## Year2002       -0.00812    0.05585   -0.15   0.884
## Year2003       -0.02657    0.05733   -0.46   0.643
## Year2004        0.02767    0.05340    0.52   0.604
## Year2005       -0.01012    0.05318   -0.19   0.849
## Year2006        0.02082    0.05224    0.40   0.690
```

```

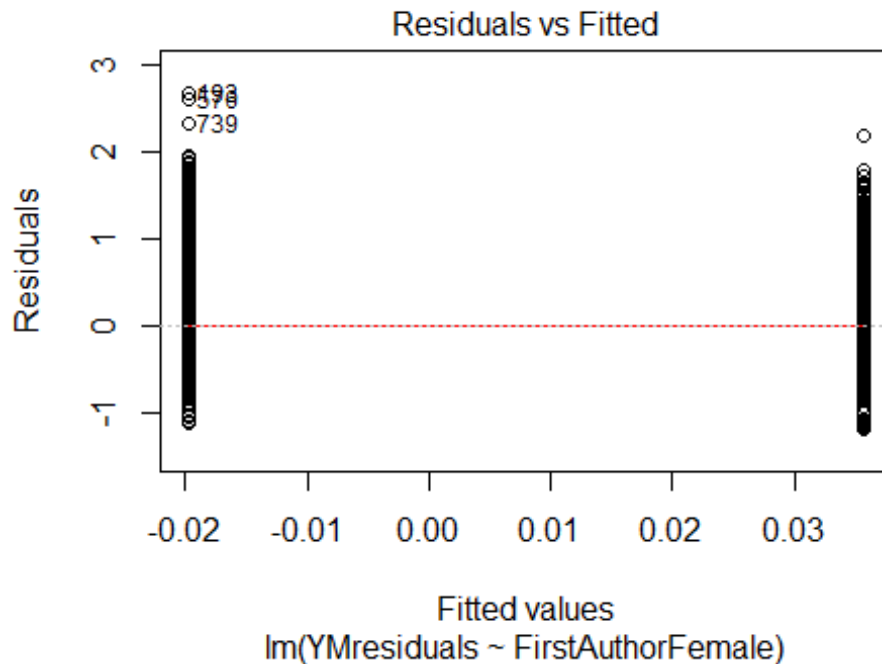
## Year2007      0.06046    0.05071    1.19    0.233
## Year2008      0.00587    0.04991    0.12    0.906
## Year2009      0.01876    0.05002    0.37    0.708
## Year2010      0.03037    0.04955    0.61    0.540
## Year2011      0.05743    0.04928    1.17    0.244
## Year2012      0.05830    0.04918    1.19    0.236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.00483,    Adjusted R-squared:  0.00302
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 746 weights are ~= 1. The remaining 8625 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.040  0.873  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      1.07e-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: 9371"
## [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
## 271 282 315 264 301 344 388 358 386 388 457 431 547 587 647
## 2011 2012
## 563 651
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 180 194 228 166 166 141 283 290 300 322 363 328 432 463 520
## 2011 2012

```

```
## 441 499
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 175 186 218 159 160 135 269 280 285 305 342 312 408 440 484
## 2011 2012
## 421 469
## [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
```

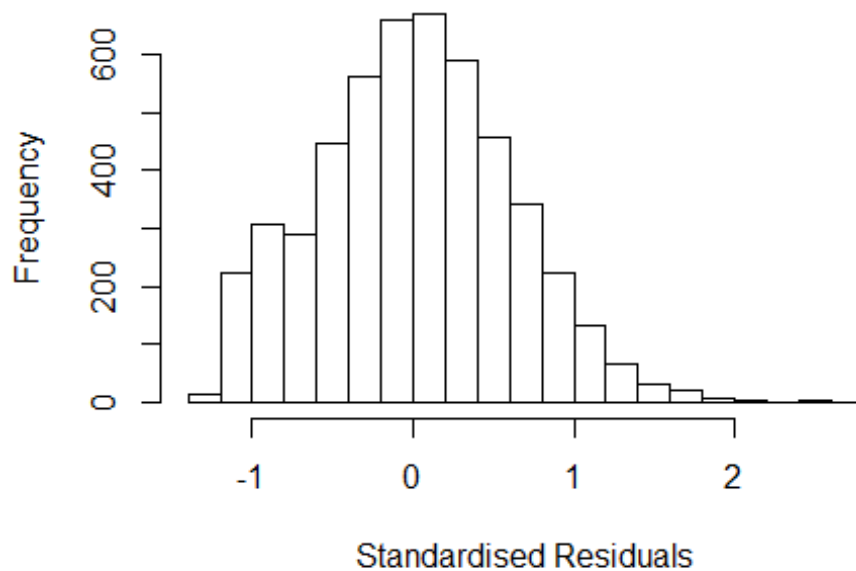


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.9, df = 1, p-value = 0.03
```

```
## [1] "Female first author team size 2018 geometric mean: 1.8344971432532"
## [1] "Male first author team size 2018 geometric mean: 1.86450698922855"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.79008967981313"
## [1] "Male last author team size 2018 geometric mean: 1.90904773583661"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 37000, 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 1.681 1 1.296
## LastAuthorFemale 1.666 1 1.291
## UniqueAuthors 1.108 4 1.013
## Year 1.151 16 1.004
```

Residuals from first and last author and team size



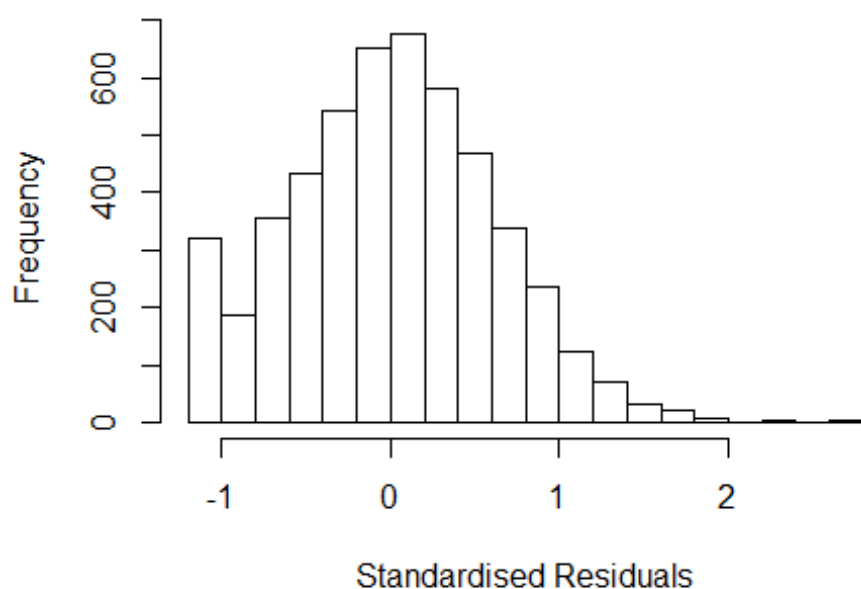
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 193 0040913897 3.627 1996      3305      2      2.602
## 576 0031435009 3.522 1997      1710      3      2.530
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.27043 -0.40828  0.00875  0.40715  2.60195
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.88028    0.05777   15.24 < 2e-16 ***
## FirstAuthorFemale1 0.05346    0.02322    2.30  0.0214 *
## LastAuthorFemale1 0.01590    0.02344    0.68  0.4974
## UniqueAuthors2    0.14477    0.02138    6.77 1.4e-11 ***
## UniqueAuthors3    0.16276    0.02951    5.52 3.6e-08 ***
## UniqueAuthors4    0.13988    0.04946    2.83  0.0047 **
## UniqueAuthors5    0.25326    0.06094    4.16 3.3e-05 ***
## Year1997        -0.05142    0.07843   -0.66  0.5121
## Year1998         0.00559    0.07745    0.07  0.9424
```

```

## Year1999      -0.05271    0.08325   -0.63    0.5267
## Year2000      -0.00113    0.08380   -0.01    0.9892
## Year2001       0.09070    0.08002    1.13    0.2571
## Year2002       0.10145    0.06950    1.46    0.1444
## Year2003       0.10094    0.06817    1.48    0.1388
## Year2004       0.09036    0.06836    1.32    0.1863
## Year2005       0.09051    0.06621    1.37    0.1717
## Year2006       0.08343    0.06521    1.28    0.2008
## Year2007       0.16586    0.06502    2.55    0.0108 *
## Year2008       0.06812    0.06595    1.03    0.3017
## Year2009       0.16054    0.06365    2.52    0.0117 *
## Year2010       0.15943    0.06304    2.53    0.0115 *
## Year2011       0.12600    0.06412    1.96    0.0495 *
## Year2012       0.07441    0.06513    1.14    0.2533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0273
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 422 weights are ~= 1. The remaining 4626 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0261 0.8700 0.9500 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      1.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.658 1      1.288
## LastAuthorFemale  1.643 1      1.282
## Year              1.051 16      1.002

```

Residuals from first and last author



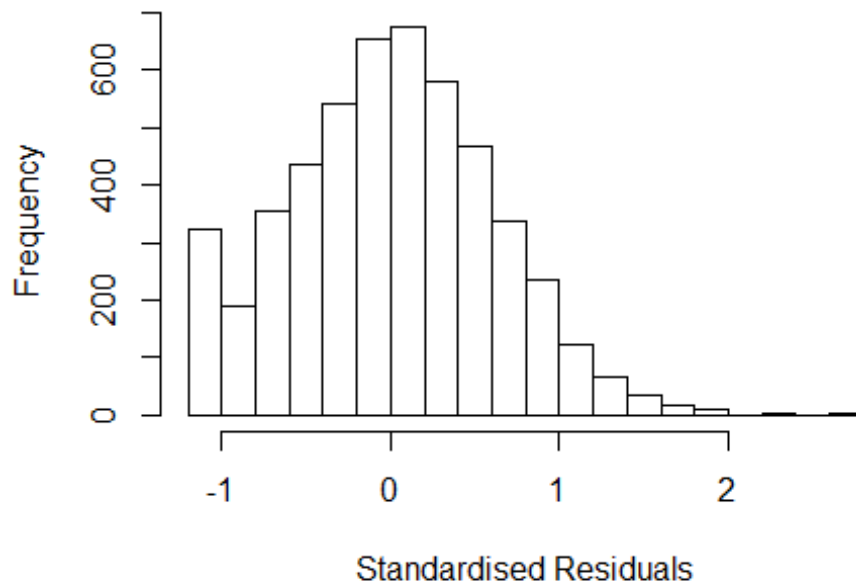
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 193 0040913897 3.627 1996      3305      2      2.707
## 576 0031435009 3.522 1997      1710      3      2.661
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1776 -0.4107  0.0107  0.4146  2.7072
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91977    0.05737   16.03  <2e-16 ***
## FirstAuthorFemale1 0.06286    0.02329    2.70  0.0070 **
## LastAuthorFemale1 0.00412    0.02352    0.18  0.8610
## Year1997       -0.05872    0.07748   -0.76  0.4485
## Year1998        0.00976    0.07790    0.13  0.9003
## Year1999       -0.05102    0.08294   -0.62  0.5385
## Year2000        0.01220    0.08421    0.14  0.8848
## Year2001        0.11919    0.08020    1.49  0.1373
## Year2002        0.11246    0.06916    1.63  0.1040
## Year2003        0.10483    0.06774    1.55  0.1218
## Year2004        0.09554    0.06823    1.40  0.1615
```

```

## Year2005          0.10892      0.06592      1.65      0.0985 .
## Year2006          0.09442      0.06496      1.45      0.1461
## Year2007          0.19453      0.06470      3.01      0.0027 **
## Year2008          0.08709      0.06558      1.33      0.1842
## Year2009          0.19087      0.06329      3.02      0.0026 **
## Year2010          0.18834      0.06263      3.01      0.0027 **
## Year2011          0.16252      0.06350      2.56      0.0105 *
## Year2012          0.10558      0.06497      1.63      0.1042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0168, Adjusted R-squared:  0.0133
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 435 weights are ~ = 1. The remaining 4613 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0139 0.8700 0.9500 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      1.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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



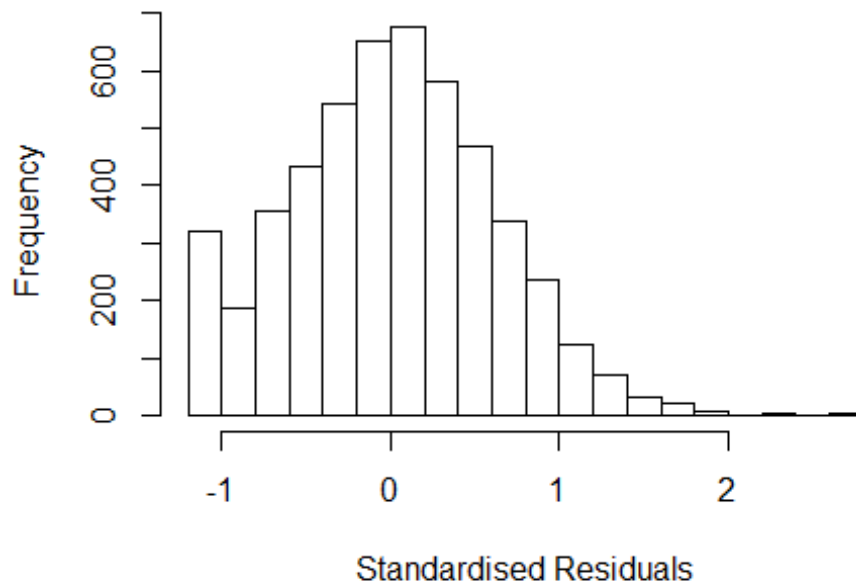
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 193 0040913897 3.627 1996      3305      2      2.707
## 576 0031435009 3.522 1997      1710      3      2.661
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1801 -0.4112  0.0111  0.4145  2.7068
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92019    0.05722   16.08 < 2e-16 ***
## FirstAuthorFemale1 0.06549    0.01845    3.55 0.00039 ***
## Year1997       -0.05881    0.07747   -0.76 0.44781
## Year1998        0.00973    0.07791    0.12 0.90059
## Year1999       -0.05112    0.08292   -0.62 0.53758
## Year2000        0.01222    0.08423    0.15 0.88461
## Year2001        0.11930    0.08022    1.49 0.13702
## Year2002        0.11250    0.06917    1.63 0.10392
## Year2003        0.10474    0.06772    1.55 0.12200
## Year2004        0.09548    0.06822    1.40 0.16170
## Year2005        0.10893    0.06592    1.65 0.09851 .
```

```

## Year2006          0.09448      0.06497      1.45  0.14596
## Year2007          0.19441      0.06468      3.01  0.00266 **
## Year2008          0.08727      0.06559      1.33  0.18341
## Year2009          0.19095      0.06331      3.02  0.00257 **
## Year2010          0.18843      0.06265      3.01  0.00264 **
## Year2011          0.16252      0.06350      2.56  0.01052 *
## Year2012          0.10570      0.06498      1.63  0.10388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0168, Adjusted R-squared:  0.0135
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 433 weights are ~= 1. The remaining 4615 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0139 0.8700 0.9500 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      1.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## 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 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 193 0040913897 3.627 1996    3305      2    2.707
## 576 0031435009 3.522 1997    1710      3    2.661
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1671 -0.4166  0.0109  0.4158  2.7021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9249    0.0572   16.17 <2e-16 ***
## LastAuthorFemale1 0.0450    0.0186    2.41  0.0158 *
## Year1997       -0.0583    0.0772   -0.75  0.4504
## Year1998        0.0102    0.0777    0.13  0.8955
## Year1999       -0.0485    0.0828   -0.59  0.5582
## Year2000        0.0111    0.0839    0.13  0.8945
## Year2001        0.1230    0.0798    1.54  0.1231
## Year2002        0.1112    0.0690    1.61  0.1074
## Year2003        0.1074    0.0676    1.59  0.1121
## Year2004        0.1008    0.0679    1.48  0.1379
## Year2005        0.1113    0.0657    1.69  0.0904 .
```

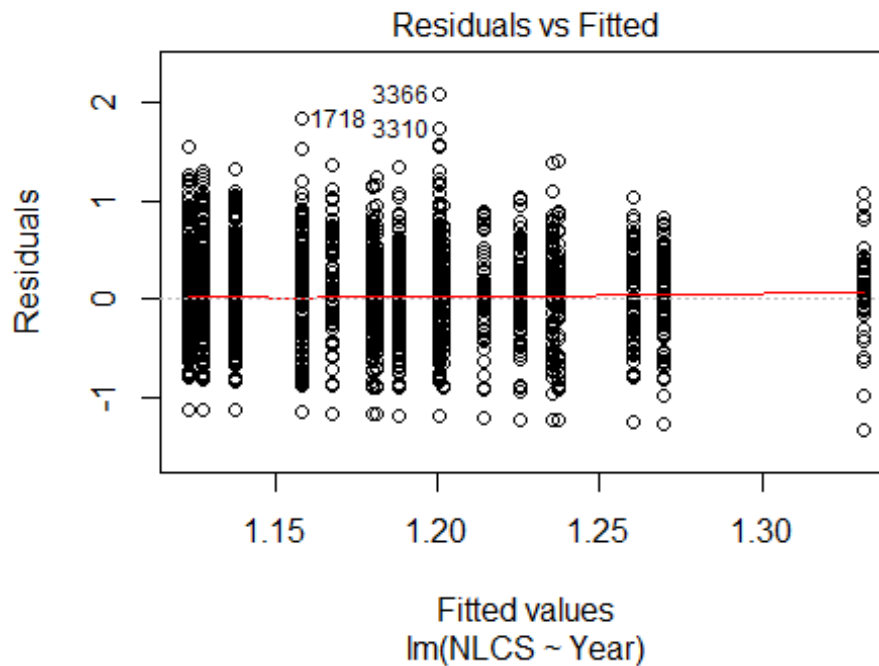


```

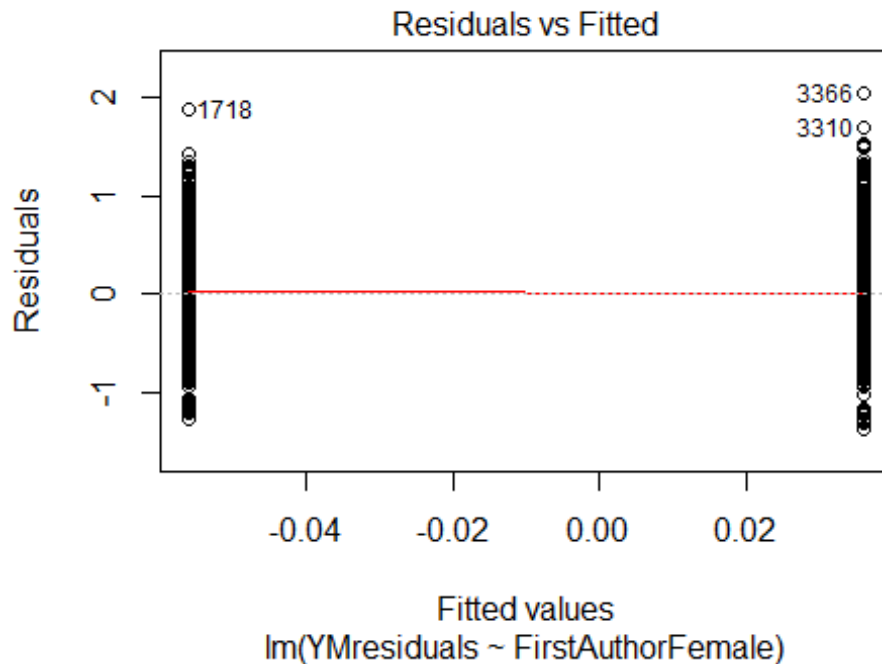
## Year2006          0.0973      0.0648      1.50      0.1333
## Year2007          0.2020      0.0645      3.13      0.0017 **
## Year2008          0.0907      0.0654      1.39      0.1654
## Year2009          0.1972      0.0631      3.13      0.0018 **
## Year2010          0.1934      0.0624      3.10      0.0020 **
## Year2011          0.1687      0.0633      2.67      0.0077 **
## Year2012          0.1103      0.0648      1.70      0.0886 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.0121
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 437 weights are ~= 1. The remaining 4611 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0149 0.8710 0.9500 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      1.98e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5048"
## [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
##   79   87  100  101  100  120  117  111  104  113  188  245  281  293  276
## 2011 2012
##  369  371
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   60   64   78   79   67   85   87   80   85   99  154  195  221  246  239

```

```
## 2011 2012
## 298 286
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 59 70 78 59 76 84 75 75 86 133 187 207 228 219
## 2011 2012
## 283 263
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 5e-04
```

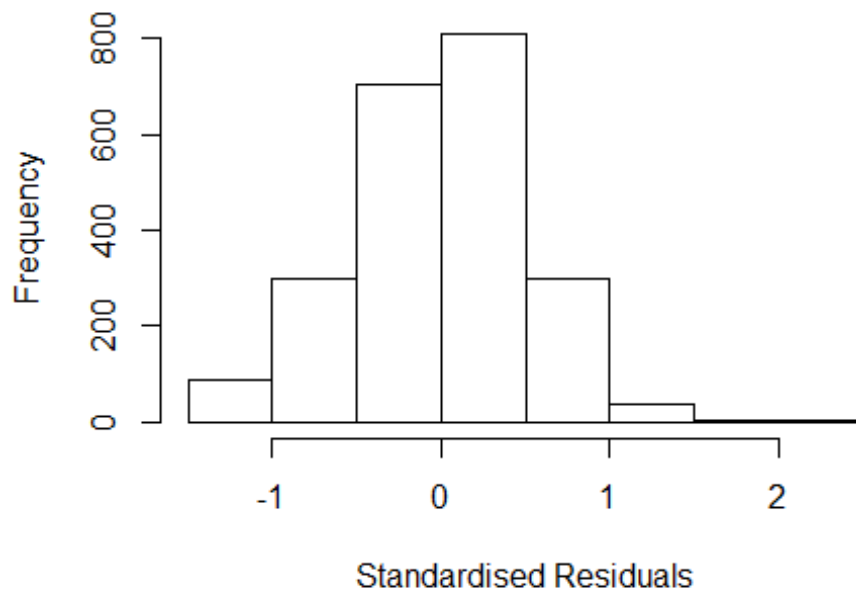


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.8, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 2.97116875560029"
## [1] "Male first author team size 2018 geometric mean: 2.55615884200206"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.76789523144363"
## [1] "Male last author team size 2018 geometric mean: 2.96727602888335"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, 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  1.160 1      1.077
## UniqueAuthors     1.203 4      1.023
## Year              1.306 16      1.008
```

Residuals from first and last author and team size



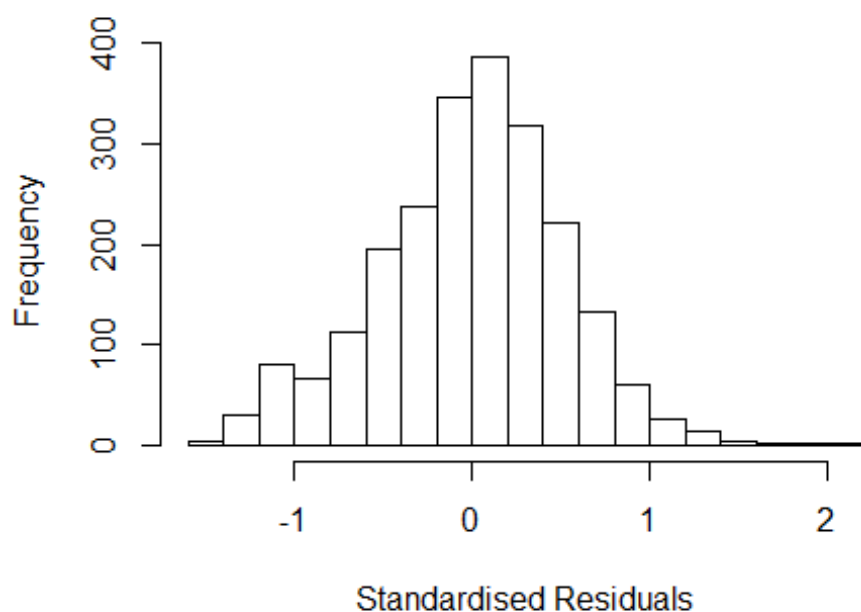
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4444 -0.3287 0.0164 0.3292 2.1959
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21974 0.06475 18.84 < 2e-16 ***
## FirstAuthorFemale1 0.07586 0.02478 3.06 0.00223 **
## LastAuthorFemale1 -0.00362 0.02352 -0.15 0.87783
## UniqueAuthors2 0.14881 0.03167 4.70 2.8e-06 ***
## UniqueAuthors3 0.17993 0.03185 5.65 1.8e-08 ***
## UniqueAuthors4 0.24682 0.03724 6.63 4.2e-11 ***
## UniqueAuthors5 0.24815 0.03488 7.11 1.5e-12 ***
## Year1997 -0.10008 0.09867 -1.01 0.31056
## Year1998 -0.08830 0.08351 -1.06 0.29048
## Year1999 -0.13415 0.07565 -1.77 0.07633 .
```

```

## Year2000      -0.17333    0.08083    -2.14    0.03211 *
## Year2001      -0.22353    0.09283    -2.41    0.01613 *
## Year2002      -0.13773    0.08527    -1.62    0.10638
## Year2003      -0.16872    0.08210    -2.06    0.03999 *
## Year2004      -0.11423    0.08808    -1.30    0.19482
## Year2005      -0.16754    0.07594    -2.21    0.02747 *
## Year2006      -0.23597    0.07405    -3.19    0.00146 **
## Year2007      -0.21010    0.06993    -3.00    0.00269 **
## Year2008      -0.23816    0.07117    -3.35    0.00083 ***
## Year2009      -0.24701    0.07071    -3.49    0.00049 ***
## Year2010      -0.27020    0.07022    -3.85    0.00012 ***
## Year2011      -0.28012    0.06989    -4.01    6.3e-05 ***
## Year2012      -0.20490    0.07131    -2.87    0.00410 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0544, Adjusted R-squared:  0.045
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 195 weights are ~= 1. The remaining 2043 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0058 0.8610 0.9490 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          4.47e-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 1.175 1 1.084
## LastAuthorFemale 1.132 1 1.064
## Year 1.104 16 1.003

```

Residuals from first and last author



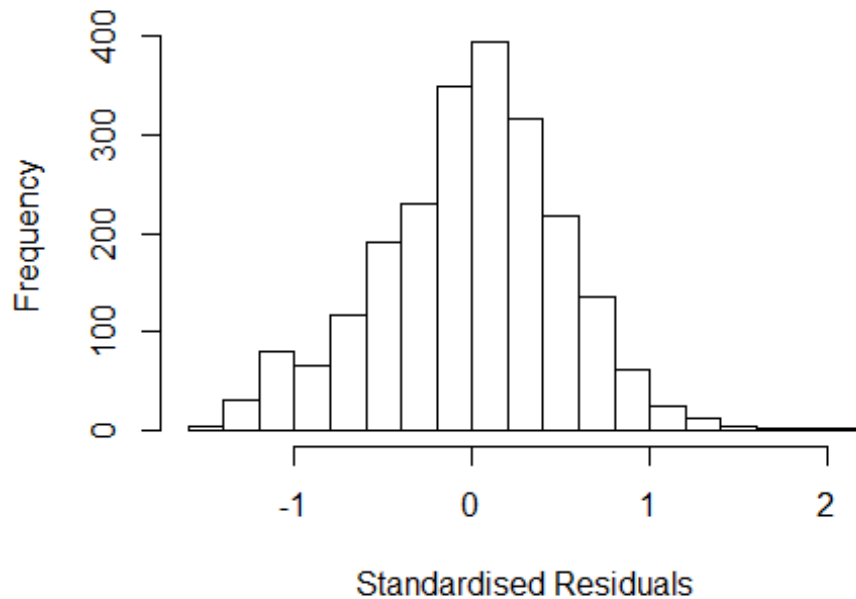
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4384 -0.3367 0.0257 0.3317 2.0448
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3406 0.0615 21.81 < 2e-16 ***
## FirstAuthorFemale1 0.0978 0.0251 3.90 0.00010 ***
## LastAuthorFemale1 -0.0181 0.0236 -0.77 0.44264
## Year1997 -0.1327 0.0973 -1.36 0.17267
## Year1998 -0.1122 0.0854 -1.31 0.18891
## Year1999 -0.1482 0.0756 -1.96 0.05026 .
## Year2000 -0.1796 0.0786 -2.29 0.02238 *
## Year2001 -0.2161 0.0925 -2.34 0.01959 *
## Year2002 -0.1328 0.0849 -1.56 0.11790
## Year2003 -0.1812 0.0815 -2.22 0.02638 *
## Year2004 -0.0921 0.0869 -1.06 0.28952
## Year2005 -0.1819 0.0755 -2.41 0.01609 *
```

```

## Year2006          -0.2130      0.0750   -2.84  0.00454 **
## Year2007          -0.2012      0.0697   -2.89  0.00391 **
## Year2008          -0.2409      0.0704   -3.42  0.00064 ***
## Year2009          -0.2515      0.0705   -3.57  0.00037 ***
## Year2010          -0.2745      0.0702   -3.91  9.4e-05 ***
## Year2011          -0.2657      0.0699   -3.80  0.00015 ***
## Year2012          -0.1821      0.0712   -2.56  0.01059 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0193, Adjusted R-squared:  0.0114
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 208 weights are ~= 1. The remaining 2030 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0438 0.8600 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      4.47e-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 1.074 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
## -1.4264 -0.3312 0.0273 0.3298 2.0391
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3360 0.0609 21.94 < 2e-16 ***
## FirstAuthorFemale1 0.0904 0.0241 3.75 0.00018 ***
## Year1997 -0.1325 0.0973 -1.36 0.17347
## Year1998 -0.1125 0.0856 -1.31 0.18884
## Year1999 -0.1481 0.0757 -1.96 0.05067 .
## Year2000 -0.1808 0.0788 -2.30 0.02179 *
## Year2001 -0.2179 0.0927 -2.35 0.01877 *
## Year2002 -0.1326 0.0850 -1.56 0.11898
## Year2003 -0.1812 0.0815 -2.22 0.02637 *
## Year2004 -0.0943 0.0867 -1.09 0.27702
## Year2005 -0.1818 0.0756 -2.41 0.01621 *
## Year2006 -0.2141 0.0750 -2.85 0.00437 **
```

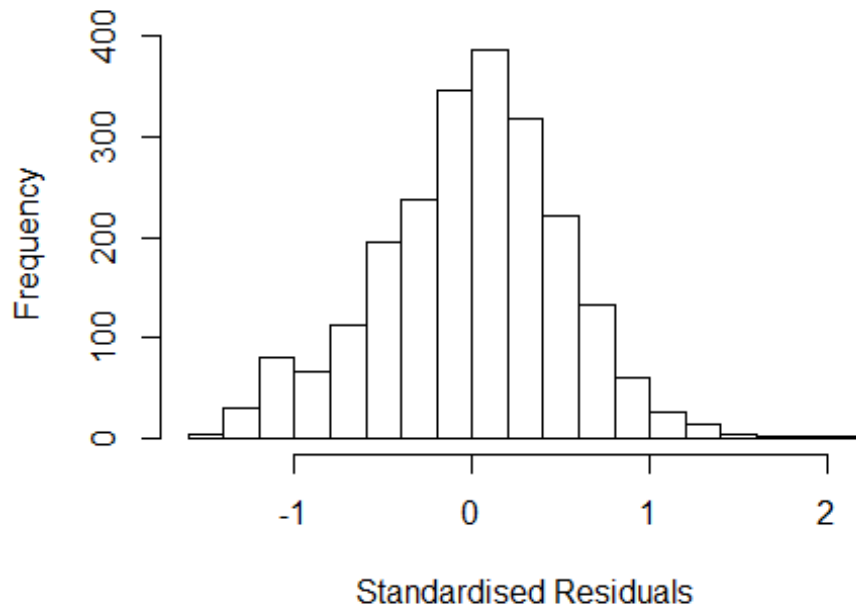


```

## Year2007          -0.2015      0.0698   -2.89   0.00392 **
## Year2008          -0.2423      0.0706   -3.43   0.00061 ***
## Year2009          -0.2522      0.0706   -3.57   0.00036 ***
## Year2010          -0.2761      0.0703   -3.93   8.8e-05 ***
## Year2011          -0.2665      0.0700   -3.81   0.00014 ***
## Year2012          -0.1825      0.0713   -2.56   0.01058 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 210 weights are ~= 1. The remaining 2028 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0456 0.8610 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      4.47e-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 1.036 1      1.018
## Year      1.036 16      1.001

```

Residuals from last author



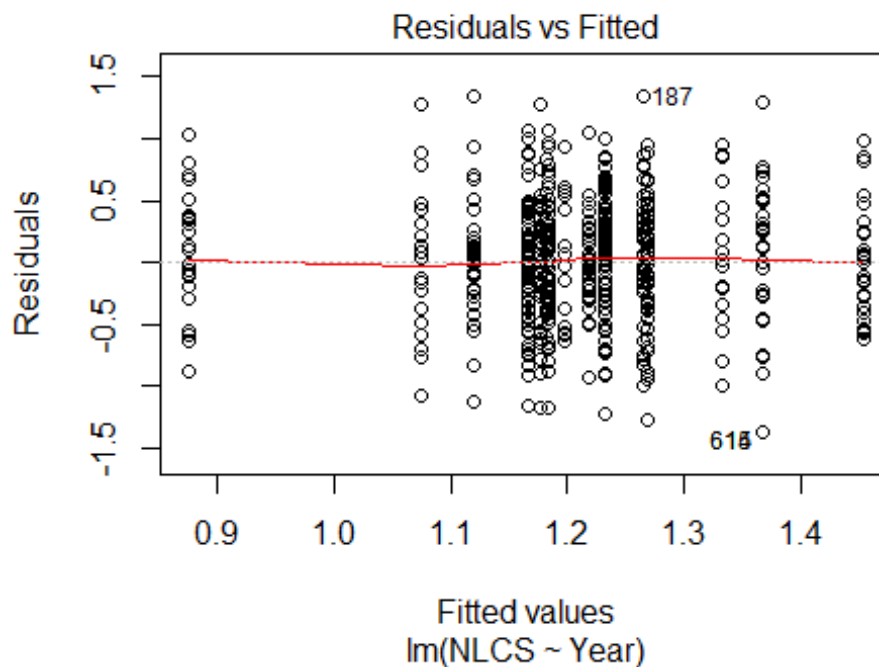
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3832 -0.3337 0.0234 0.3311 2.0602
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3636 0.0626 21.79 < 2e-16 ***
## LastAuthorFemale1 0.0196 0.0228 0.86 0.39161
## Year1997 -0.1182 0.0973 -1.21 0.22471
## Year1998 -0.1032 0.0860 -1.20 0.23028
## Year1999 -0.1418 0.0768 -1.84 0.06522 .
## Year2000 -0.1743 0.0803 -2.17 0.03005 *
## Year2001 -0.2053 0.0934 -2.20 0.02802 *
## Year2002 -0.1217 0.0857 -1.42 0.15561
## Year2003 -0.1700 0.0820 -2.07 0.03818 *
## Year2004 -0.0843 0.0884 -0.95 0.34014
## Year2005 -0.1669 0.0767 -2.17 0.02974 *
## Year2006 -0.1902 0.0753 -2.53 0.01157 *
```

```

## Year2007          -0.1763      0.0703    -2.51  0.01227 *
## Year2008          -0.2231      0.0716    -3.12  0.00186 **
## Year2009          -0.2320      0.0712    -3.26  0.00114 **
## Year2010          -0.2546      0.0708    -3.60  0.00033 ***
## Year2011          -0.2477      0.0706    -3.51  0.00046 ***
## Year2012          -0.1604      0.0718    -2.23  0.02563 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0121, Adjusted R-squared:  0.00449
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 207 weights are ~= 1. The remaining 2031 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0387 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      4.47e-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: 2238"
## [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
##   34   26   33   17   52   42   30   30   45   25   44   42   63   44   66
## 2011 2012
##   82   74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   18   28   15   35   28   18   24   40   20   32   35   49   31   55
## 2011 2012

```

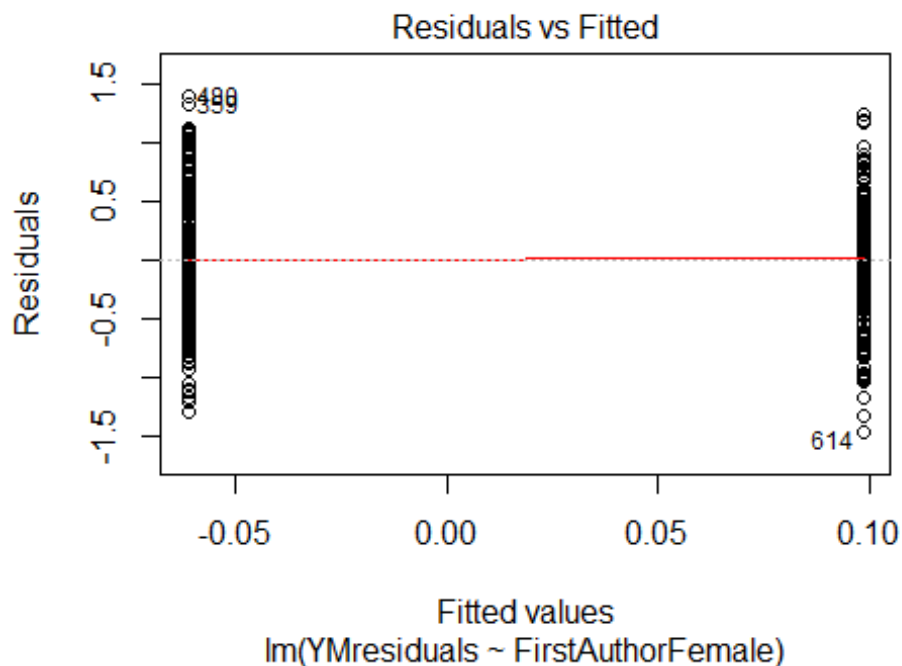
```
## 67 55
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 16 24 14 31 24 18 23 35 19 30 34 40 29 54
## 2011 2012
## 63 53
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.06
```



```
##
## 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: 3.59120688484746"
## [1] "Male first author team size 2018 geometric mean: 2.8552192217299"
## 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 = 600, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.17991562118665"
## [1] "Male last author team size 2018 geometric mean: 3.11354242318146"

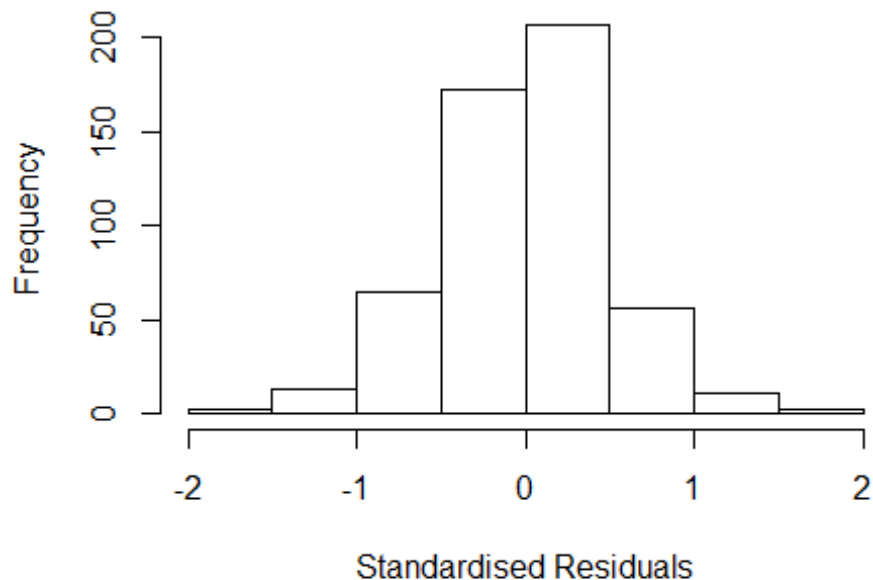
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## 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	1.286	1	1.134
LastAuthorFemale	1.113	1	1.055
UniqueAuthors	2.067	4	1.095
Year	2.388	16	1.028

Residuals from first and last author and team size



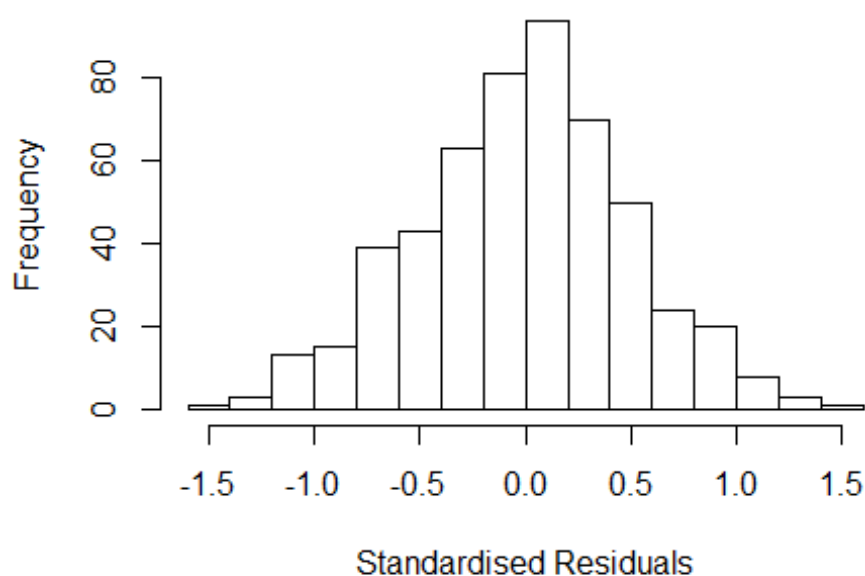
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.729 -0.317  0.023  0.282  1.647
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9255    0.1352   6.84 2.2e-11 ***
## FirstAuthorFemale1 0.1022    0.0460   2.22 0.02691 *
## LastAuthorFemale1 0.0801    0.0434   1.85 0.06560 .
## UniqueAuthors2    0.2165    0.0604   3.59 0.00037 ***
## UniqueAuthors3    0.3817    0.0638   5.98 4.2e-09 ***
## UniqueAuthors4    0.4294    0.0753   5.70 2.0e-08 ***
## UniqueAuthors5    0.4929    0.0837   5.89 7.0e-09 ***
## Year1997          0.2261    0.1959   1.15 0.24905
## Year1998         -0.2916    0.1782  -1.64 0.10239
## Year1999          0.1097    0.1740   0.63 0.52862
```

```

## Year2000          0.0171      0.1697      0.10  0.91975
## Year2001          0.2949      0.1733      1.70  0.08944 .
## Year2002          0.0295      0.1816      0.16  0.87099
## Year2003         -0.0317      0.1613     -0.20  0.84432
## Year2004         -0.0720      0.1622     -0.44  0.65736
## Year2005         -0.1139      0.1983     -0.57  0.56608
## Year2006          0.0753      0.1512      0.50  0.61895
## Year2007         -0.1110      0.1604     -0.69  0.48909
## Year2008         -0.1005      0.1536     -0.65  0.51323
## Year2009          0.1281      0.1720      0.74  0.45665
## Year2010         -0.0961      0.1555     -0.62  0.53704
## Year2011         -0.0369      0.1633     -0.23  0.82150
## Year2012         -0.0647      0.1588     -0.41  0.68369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.194, Adjusted R-squared:  0.159
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 484 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.120  0.870  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      1.89e-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 1.264 1      1.124
## LastAuthorFemale  1.114 1      1.055
## Year              1.336 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
## -1.530 -0.343  0.013  0.325  1.442
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9970    0.1383   7.21 2e-12 ***
## FirstAuthorFemale1 0.1517    0.0492   3.08 0.0022 **
## LastAuthorFemale1  0.0767    0.0475   1.61 0.1069
## Year1997          0.2949    0.2140   1.38 0.1688
## Year1998         -0.2174    0.1878  -1.16 0.2478
## Year1999          0.1696    0.2002   0.85 0.3973
## Year2000          0.1319    0.1819   0.73 0.4686
## Year2001          0.3846    0.1722   2.23 0.0260 *
## Year2002          0.1401    0.1848   0.76 0.4487
## Year2003          0.1176    0.1590   0.74 0.4598
## Year2004          0.1015    0.1592   0.64 0.5238
## Year2005          0.0360    0.1922   0.19 0.8516
```

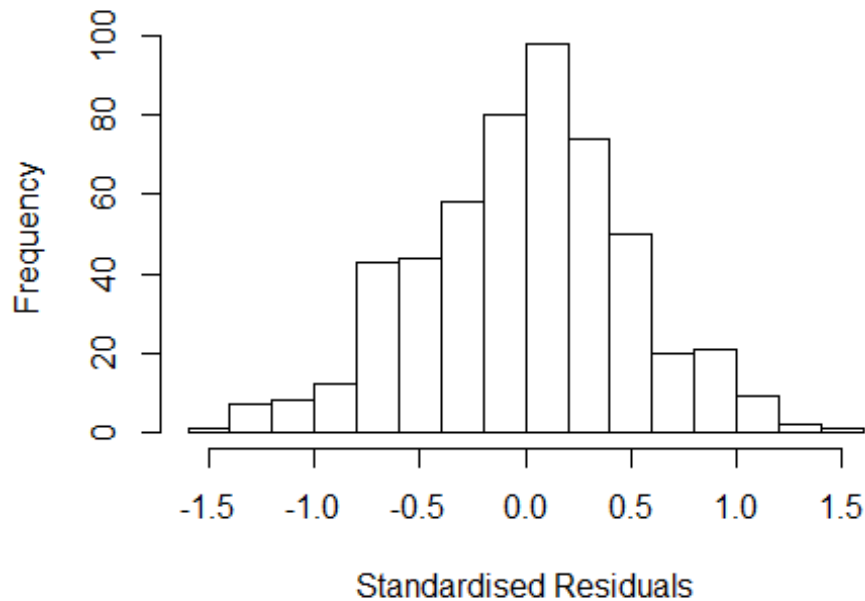


```

## Year2006          0.1590      0.1539      1.03      0.3022
## Year2007          0.0219      0.1594      0.14      0.8909
## Year2008          0.1001      0.1561      0.64      0.5216
## Year2009          0.3051      0.1854      1.65      0.1005
## Year2010          0.0946      0.1555      0.61      0.5434
## Year2011          0.1964      0.1655      1.19      0.2359
## Year2012          0.1649      0.1605      1.03      0.3048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.084, Adjusted R-squared:  0.0516
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 481 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.301  0.858  0.950  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      1.89e-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 1.242 1          1.114
## Year              1.242 16          1.007

```

Residuals from first author



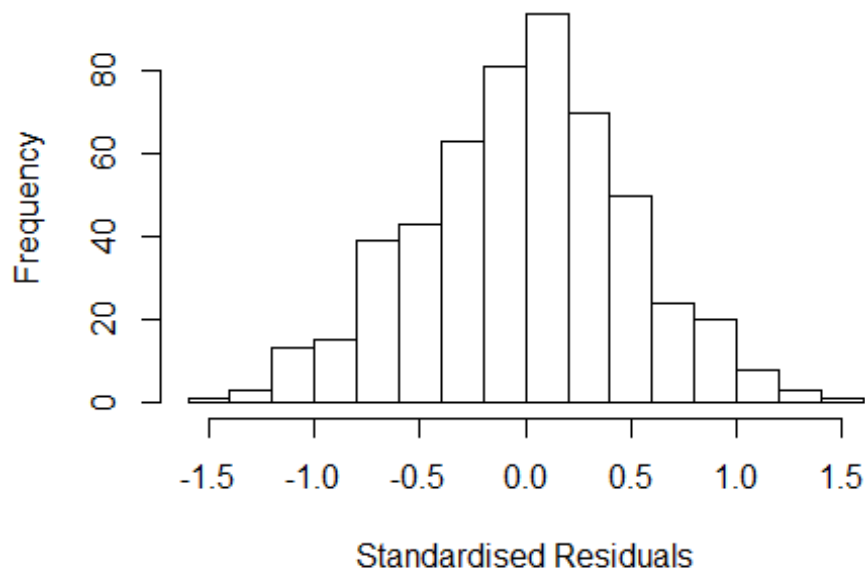
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4931 -0.3383  0.0261  0.3242  1.4235
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0054     0.1398   7.19 2.3e-12 ***
## FirstAuthorFemale1  0.1745     0.0489   3.57  0.0004 ***
## Year1997          0.2951     0.2200   1.34  0.1804
## Year1998         -0.2131     0.1897  -1.12  0.2619
## Year1999          0.1714     0.2050   0.84  0.4034
## Year2000          0.1457     0.1826   0.80  0.4254
## Year2001          0.3913     0.1737   2.25  0.0247 *
## Year2002          0.1511     0.1862   0.81  0.4175
## Year2003          0.1292     0.1589   0.81  0.4166
## Year2004          0.1138     0.1603   0.71  0.4781
## Year2005          0.0522     0.1908   0.27  0.7846
## Year2006          0.1683     0.1547   1.09  0.2771
```

```

## Year2007          0.0321      0.1603      0.20      0.8414
## Year2008          0.1136      0.1562      0.73      0.4675
## Year2009          0.3132      0.1871      1.67      0.0948 .
## Year2010          0.1048      0.1570      0.67      0.5049
## Year2011          0.2058      0.1665      1.24      0.2172
## Year2012          0.1653      0.1616      1.02      0.3068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.0794, Adjusted R-squared:  0.0487
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 478 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.324  0.861  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      1.89e-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 1.091 1          1.044
## Year            1.091 16          1.003

```

Residuals from last author



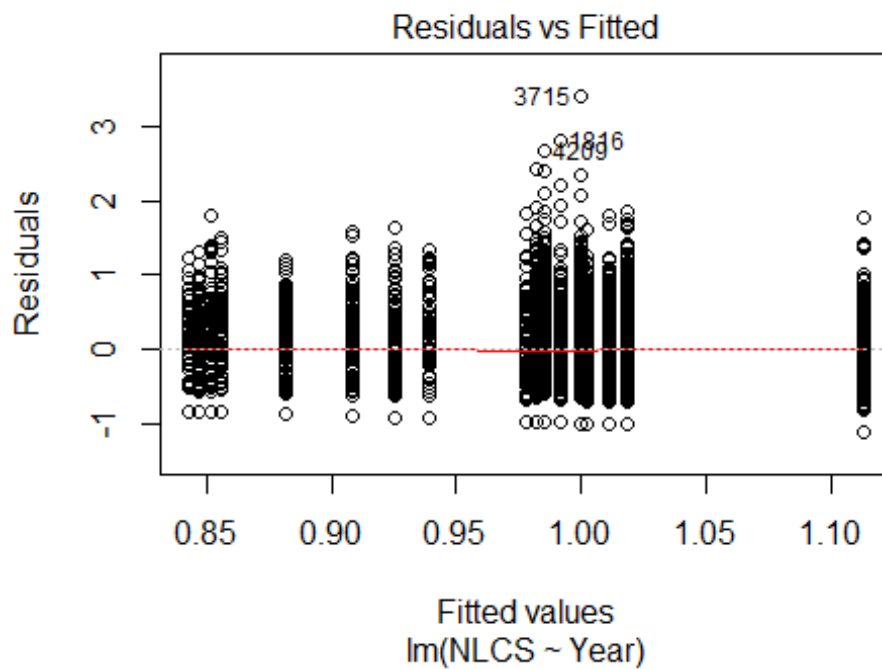
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4852 -0.3619 0.0308 0.3156 1.3917
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0142 0.1368 7.41 5.2e-13 ***
## LastAuthorFemale1 0.1214 0.0469 2.59 0.010 **
## Year1997 0.2735 0.2129 1.28 0.200
## Year1998 -0.2145 0.1871 -1.15 0.252
## Year1999 0.1828 0.1984 0.92 0.357
## Year2000 0.1468 0.1811 0.81 0.418
## Year2001 0.4115 0.1690 2.43 0.015 *
## Year2002 0.1606 0.1758 0.91 0.361
## Year2003 0.1311 0.1566 0.84 0.403
## Year2004 0.1311 0.1564 0.84 0.403
## Year2005 0.0457 0.1900 0.24 0.810
## Year2006 0.1776 0.1521 1.17 0.244
```

```

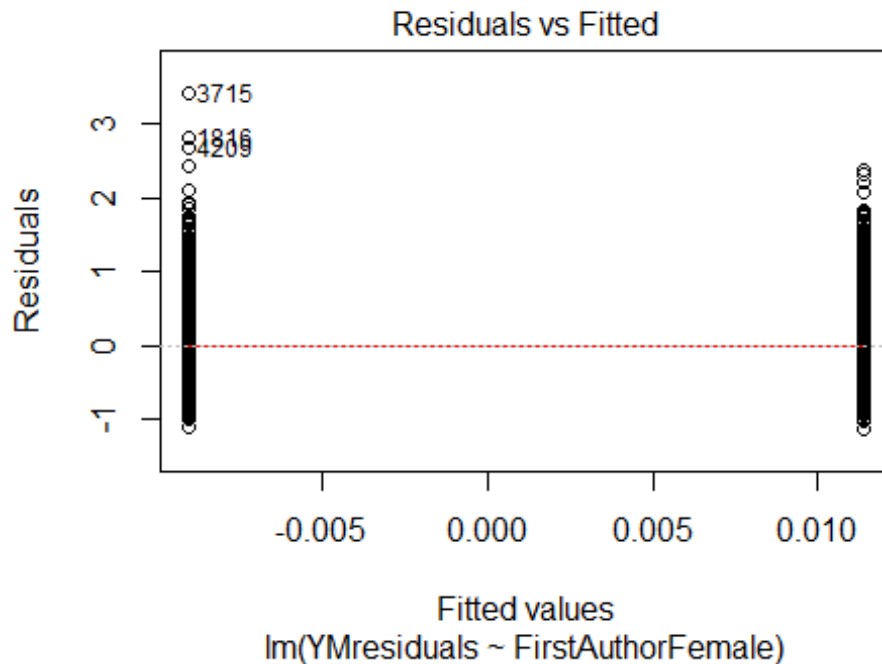
## Year2007          0.0551      0.1584      0.35      0.728
## Year2008          0.1266      0.1559      0.81      0.417
## Year2009          0.3496      0.1853      1.89      0.060 .
## Year2010          0.1159      0.1548      0.75      0.454
## Year2011          0.2370      0.1635      1.45      0.148
## Year2012          0.2389      0.1563      1.53      0.127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0688, Adjusted R-squared:  0.0377
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 488 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.331  0.857  0.948  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      1.89e-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: 528"
## [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
## 120 137 116 125 164 163 202 170 174 149 224 219 296 449 444
## 2011 2012
## 492 512
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 117 69 76 108 129 137 153 153 134 194 179 256 372 383
## 2011 2012

```

```
## 431 446
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 103 114 68 74 103 118 131 149 148 131 182 174 240 366 375
## 2011 2012
## 418 435
## [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
```

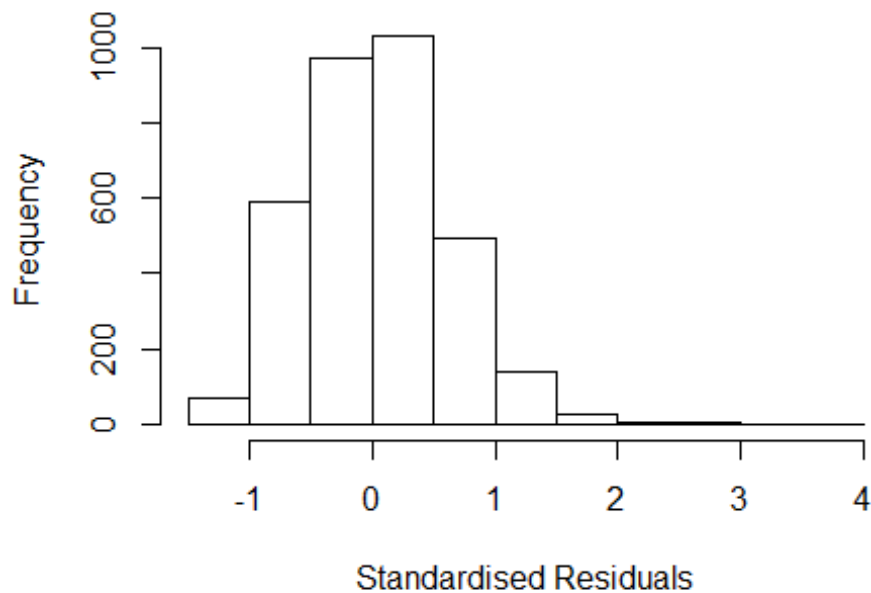


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.85, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 1.77525605183683"
## [1] "Male first author team size 2018 geometric mean: 1.53108715612525"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 29000, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.62646990363534"
## [1] "Male last author team size 2018 geometric mean: 1.68918104374499"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 25000, 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 1.986 1          1.409
## LastAuthorFemale  1.972 1          1.404
## UniqueAuthors    1.096 4          1.011
## Year              1.117 16         1.003
```

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
## 1816 31744446982 3.808 2006      1200      2      2.965
## 3320 77953476088 3.413 2010      1200      2      2.560
## 3715 79952817615 4.404 2011      1200      2      3.542
## 4209 84860556355 3.657 2012      1200      2      2.610
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4545 -0.4089  0.0193  0.4081  3.5421
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77474    0.05889   13.15 < 2e-16 ***
## FirstAuthorFemale1 0.02348    0.02993    0.78  0.4328
## LastAuthorFemale1 -0.03075    0.02999   -1.03  0.3054
## UniqueAuthors2    0.22960    0.02640    8.70 < 2e-16 ***
## UniqueAuthors3    0.41336    0.03834   10.78 < 2e-16 ***
## UniqueAuthors4    0.45601    0.05807    7.85 5.5e-15 ***
## UniqueAuthors5    0.56914    0.06856    8.30 < 2e-16 ***
```

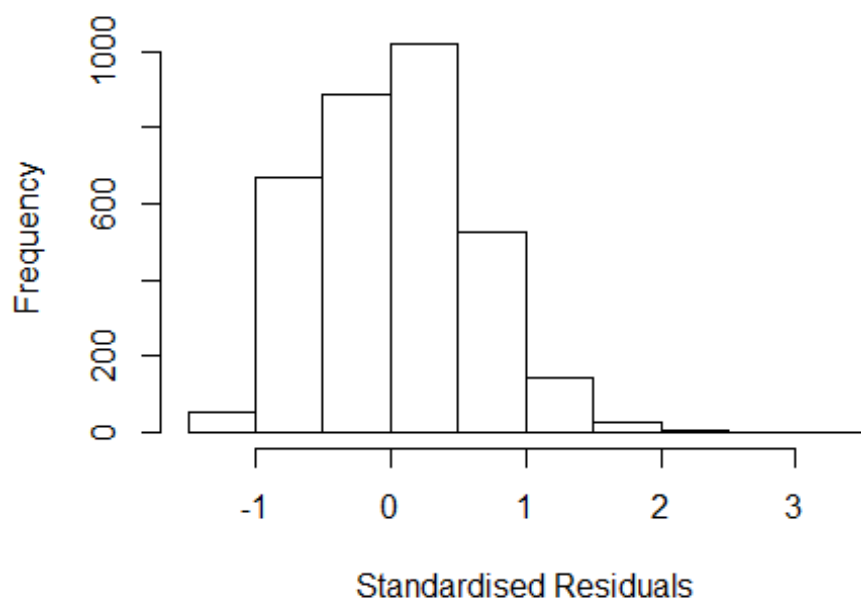


```

## Year1997      0.05901      0.08007      0.74      0.4612
## Year1998     -0.00892      0.09306     -0.10      0.9236
## Year1999      0.09276      0.09438      0.98      0.3257
## Year2000     -0.01386      0.07950     -0.17      0.8616
## Year2001      0.09686      0.08232      1.18      0.2394
## Year2002      0.12627      0.08163      1.55      0.1220
## Year2003      0.00505      0.07354      0.07      0.9452
## Year2004      0.02710      0.07768      0.35      0.7272
## Year2005     -0.04216      0.07708     -0.55      0.5844
## Year2006      0.06783      0.07246      0.94      0.3493
## Year2007      0.20536      0.07356      2.79      0.0053 **
## Year2008      0.12254      0.06893      1.78      0.0755 .
## Year2009      0.12301      0.06691      1.84      0.0661 .
## Year2010      0.07818      0.06665      1.17      0.2409
## Year2011      0.08712      0.06685      1.30      0.1926
## Year2012      0.04249      0.06635      0.64      0.5220
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0772, Adjusted R-squared:  0.0711
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1287,2692) are outliers with |weight| = 0 ( < 3e-05);
## 275 weights are ~= 1. The remaining 3052 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0219 0.8550 0.9530 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      3.00e-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 2.144 1      1.464
## LastAuthorFemale 2.132 1      1.460
## Year      1.028 16      1.001

```

Residuals from first and last author

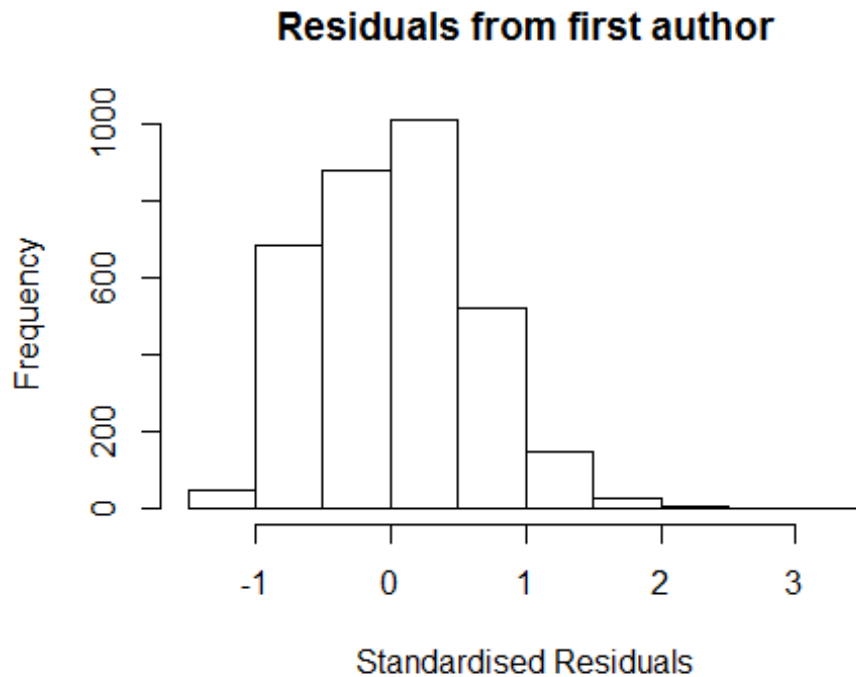


```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1816 31744446982 3.808 2006      1200      2      2.861
## 3715 79952817615 4.404 2011      1200      2      3.431
## 4209 84860556355 3.657 2012      1200      2      2.708
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1146 -0.4391  0.0186  0.4226  3.4306
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82935    0.06272   13.22 < 2e-16 ***
## FirstAuthorFemale1  0.05945    0.03243    1.83  0.06691 .
## LastAuthorFemale1 -0.05416    0.03259   -1.66  0.09660 .
## Year1997         0.07249    0.08568    0.85  0.39755
## Year1998        -0.00627    0.09622   -0.07  0.94803
## Year1999         0.12017    0.10108    1.19  0.23460
## Year2000         0.02192    0.08423    0.26  0.79472
## Year2001         0.12300    0.08613    1.43  0.15339
## Year2002         0.15558    0.08316    1.87  0.06146 .
## Year2003         0.03564    0.07698    0.46  0.64343
```

```

## Year2004          0.06529      0.08168      0.80  0.42416
## Year2005         -0.00248      0.08014     -0.03  0.97532
## Year2006          0.11771      0.07581      1.55  0.12060
## Year2007          0.27999      0.07598      3.69  0.00023 ***
## Year2008          0.17344      0.07224      2.40  0.01641 *
## Year2009          0.16301      0.07056      2.31  0.02093 *
## Year2010          0.13424      0.07065      1.90  0.05750 .
## Year2011          0.14403      0.07084      2.03  0.04211 *
## Year2012          0.11960      0.07043      1.70  0.08957 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.632
## Multiple R-squared:  0.0121, Adjusted R-squared:  0.00672
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 2692 is an outlier with |weight| = 0 ( < 3e-05);
## 292 weights are ~= 1. The remaining 3036 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0043  0.8510  0.9490  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          3.00e-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 1.017 1          1.008
## Year              1.017 16          1.001

```

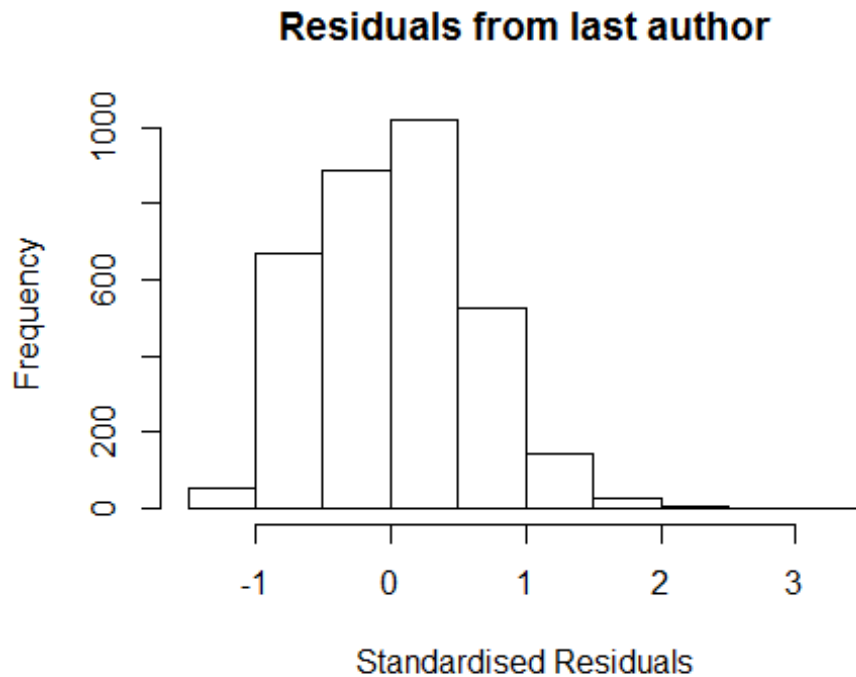


```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1816 31744446982 3.808 2006      1200      2      2.861
## 3715 79952817615 4.404 2011      1200      2      3.431
## 4209 84860556355 3.657 2012      1200      2      2.708
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1274 -0.4329  0.0192  0.4242  3.4369
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82078    0.06256   13.12 < 2e-16 ***
## FirstAuthorFemale1 0.02151    0.02238    0.96  0.33663
## Year1997        0.07504    0.08569    0.88  0.38128
## Year1998       -0.00323    0.09596   -0.03  0.97318
## Year1999        0.12110    0.10095    1.20  0.23037
## Year2000        0.02331    0.08428    0.28  0.78217
## Year2001        0.12448    0.08628    1.44  0.14917
## Year2002        0.15856    0.08322    1.91  0.05682 .
## Year2003        0.03760    0.07696    0.49  0.62523
## Year2004        0.07021    0.08181    0.86  0.39085
```

```

## Year2005          -0.00192    0.07998   -0.02  0.98083
## Year2006          0.12096    0.07579    1.60  0.11059
## Year2007          0.28513    0.07611    3.75  0.00018 ***
## Year2008          0.17579    0.07221    2.43  0.01496 *
## Year2009          0.16600    0.07056    2.35  0.01869 *
## Year2010          0.13555    0.07067    1.92  0.05520 .
## Year2011          0.14628    0.07080    2.07  0.03889 *
## Year2012          0.12321    0.07040    1.75  0.08019 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.0111, Adjusted R-squared:  0.00605
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 2692 is an outlier with |weight| = 0 ( < 3e-05);
## 290 weights are ~= 1. The remaining 3038 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0044 0.8540 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      3.00e-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 1.011 1          1.005
## Year            1.011 16          1.000

```



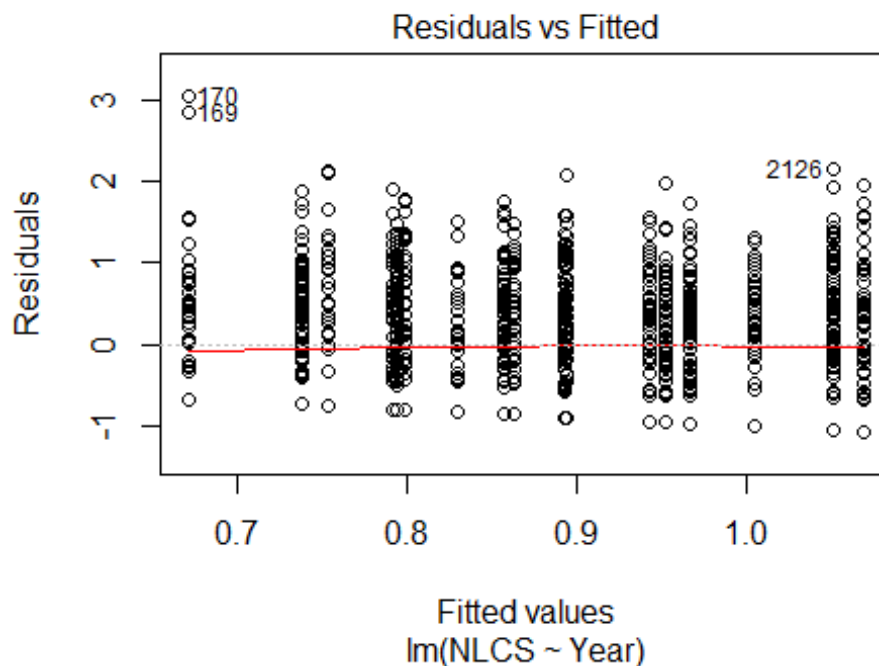
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1816 31744446982 3.808 2006      1200      2      2.861
## 3715 79952817615 4.404 2011      1200      2      3.431
## 4209 84860556355 3.657 2012      1200      2      2.708
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1211 -0.4333  0.0165  0.4217  3.4214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.835714   0.062550   13.36 < 2e-16 ***
## LastAuthorFemale1 -0.012160   0.022428   -0.54  0.58772
## Year1997         0.073075   0.085588    0.85  0.39328
## Year1998        -0.005074   0.095991   -0.05  0.95784
## Year1999         0.119977   0.100912    1.19  0.23455
## Year2000         0.021888   0.083975    0.26  0.79438
## Year2001         0.123476   0.086349    1.43  0.15282
## Year2002         0.154799   0.083095    1.86  0.06256 .
## Year2003         0.036310   0.076800    0.47  0.63639
## Year2004         0.070604   0.081583    0.87  0.38687
```

```

## Year2005          -0.000557    0.079943    -0.01    0.99444
## Year2006           0.118825    0.075652     1.57    0.11635
## Year2007           0.285355    0.076093     3.75    0.00018 ***
## Year2008           0.175707    0.072093     2.44    0.01485 *
## Year2009           0.165742    0.070470     2.35    0.01873 *
## Year2010           0.135485    0.070579     1.92    0.05499 .
## Year2011           0.146929    0.070680     2.08    0.03771 *
## Year2012           0.124735    0.070245     1.78    0.07587 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.632
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.00585
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 2692 is an outlier with |weight| = 0 ( < 3e-05);
## 281 weights are ~ = 1. The remaining 3047 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0051 0.8530 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          3.00e-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: 3329"
## [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
##   60  109   76   90   79   93   88   92   86  106  115  140  129  158  148
## 2011 2012
##   167  154
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    50   84   61   64   64   76   75   79   74   88   94  117  110  140  119
## 2011 2012
##   145  130
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    46   82   59   61   61   73   75   76   69   84   87  111  105  134  111
## 2011 2012
##   140  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 = 24, df = 16, p-value = 0.09
```

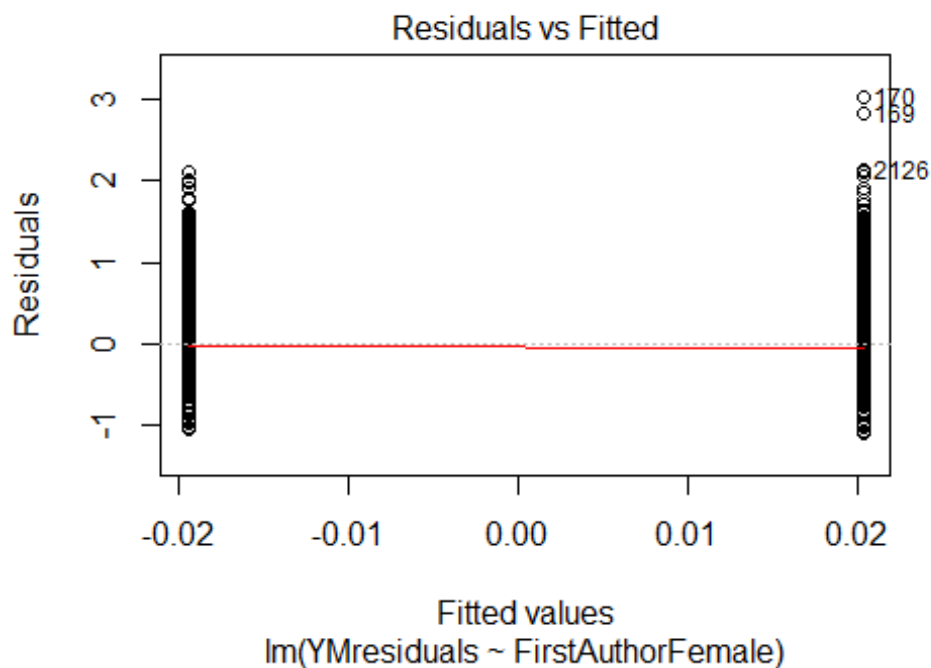


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4, df = 1, p-value = 0.05
##
## [1] "Female first author team size 2018 geometric mean: 1.53163366196329"
## [1] "Male first author team size 2018 geometric mean: 2.08681422898329"
##
## Wilcoxon rank sum test with continuity correction
```



```
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 710, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.45029772532531"
## [1] "Male last author team size 2018 geometric mean: 2.10095259485258"

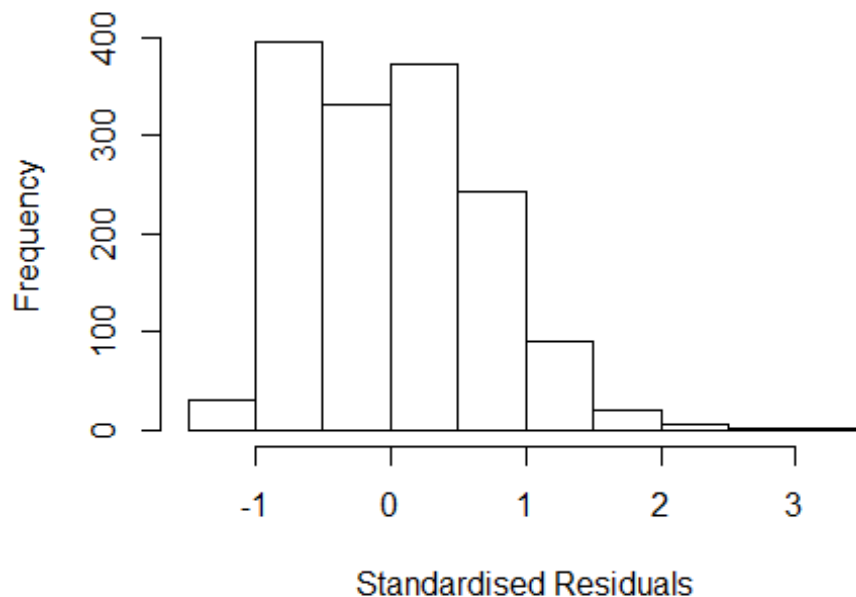
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 650, p-value = 0.005
## 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	2.529	1	1.590
LastAuthorFemale	2.442	1	1.563
UniqueAuthors	1.229	4	1.026
Year	1.302	16	1.008

Residuals from first and last author and team size



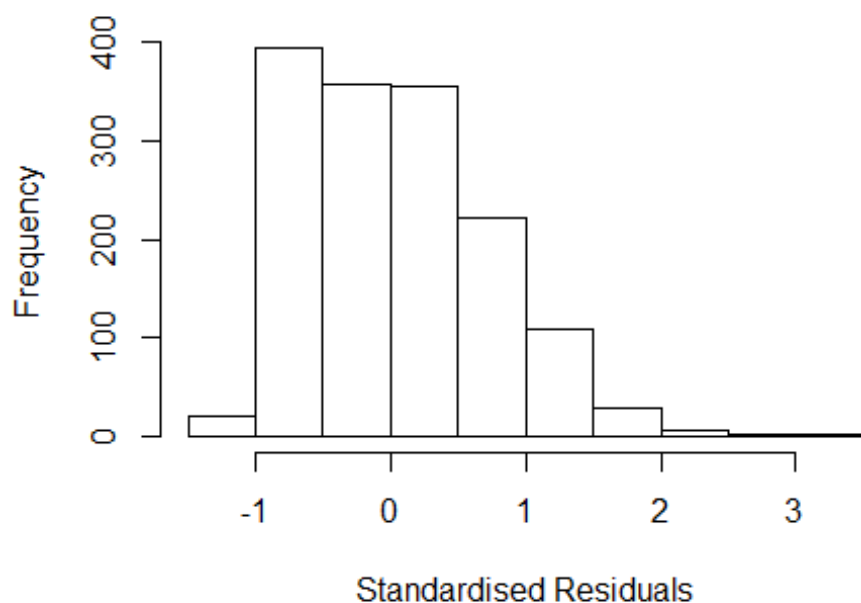
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 169 0031435009 3.522 1997    1710     3    2.620
## 170 0031482189 3.717 1997    1705     4    3.178
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4741 -0.5392 -0.0108  0.4834  3.1778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.72273    0.09462   7.64 4.0e-14 ***
## FirstAuthorFemale1 -0.02367    0.05523  -0.43  0.6684
## LastAuthorFemale1 -0.00153    0.05428  -0.03  0.9775
## UniqueAuthors2     0.37815    0.04673   8.09 1.2e-15 ***
## UniqueAuthors3     0.36324    0.06915   5.25 1.7e-07 ***
## UniqueAuthors4     0.67818    0.09549   7.10 1.9e-12 ***
## UniqueAuthors5     0.62887    0.19593   3.21  0.0014 **
## Year1997        -0.18358    0.11852  -1.55  0.1216
## Year1998        -0.15704    0.13551  -1.16  0.2467
```

```

## Year1999      -0.02367    0.13313   -0.18    0.8589
## Year2000      -0.08249    0.13492   -0.61    0.5410
## Year2001       0.18966    0.13656    1.39    0.1651
## Year2002       0.04425    0.13074    0.34    0.7351
## Year2003      -0.00601    0.11868   -0.05    0.9596
## Year2004       0.14769    0.12301    1.20    0.2301
## Year2005       0.04484    0.12351    0.36    0.7166
## Year2006       0.05731    0.11363    0.50    0.6141
## Year2007      -0.12625    0.11350   -1.11    0.2662
## Year2008       0.05431    0.11365    0.48    0.6328
## Year2009      -0.16514    0.10889   -1.52    0.1296
## Year2010      -0.03435    0.11208   -0.31    0.7593
## Year2011      -0.03278    0.10939   -0.30    0.7645
## Year2012       0.05876    0.12091    0.49    0.6271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.681
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.0967
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 125 is an outlier with |weight| <= 4.5e-05 ( < 6.7e-05);
## 137 weights are ~= 1. The remaining 1357 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.891  0.940   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           6.69e-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 2.479 1 1.574
## LastAuthorFemale 2.432 1 1.560
## Year 1.110 16 1.003

```

Residuals from first and last author



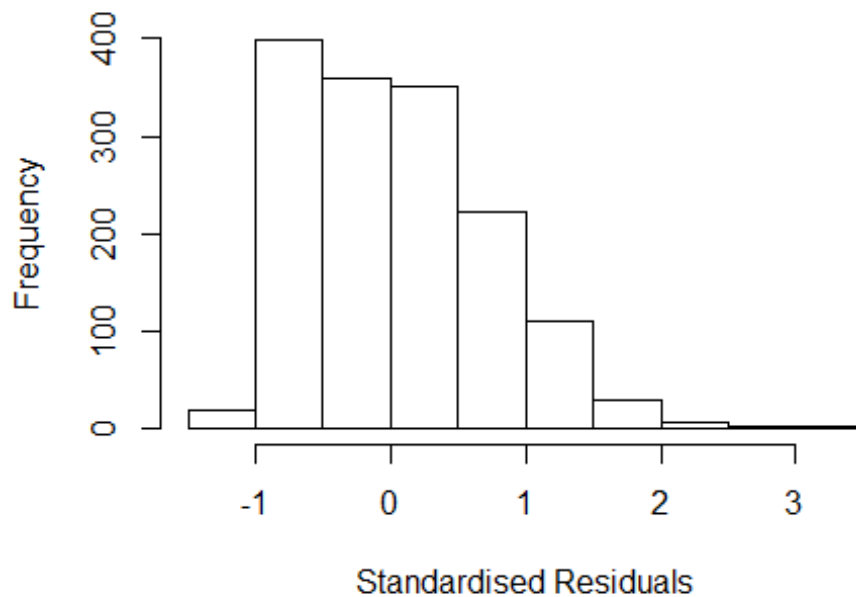
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 169 0031435009 3.522 1997    1710      3      2.935
## 170 0031482189 3.717 1997    1705      4      3.130
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0344 -0.5606 -0.0216  0.4929  3.1302
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82831    0.10026   8.26 3.2e-16 ***
## FirstAuthorFemale1 0.01905    0.05770   0.33  0.741
## LastAuthorFemale1 -0.03581    0.05696  -0.63  0.530
## Year1997        -0.24153    0.12456  -1.94  0.053 .
## Year1998        -0.17460    0.14454  -1.21  0.227
## Year1999        -0.09408    0.13629  -0.69  0.490
## Year2000        -0.11001    0.14077  -0.78  0.435
## Year2001         0.20609    0.14154   1.46  0.146
## Year2002         0.03553    0.13403   0.27  0.791
## Year2003         0.00534    0.12413   0.04  0.966
## Year2004         0.17837    0.13028   1.37  0.171
```

```

## Year2005          0.07973      0.12740      0.63      0.531
## Year2006          0.11588      0.11822      0.98      0.327
## Year2007         -0.08139      0.11997     -0.68      0.498
## Year2008          0.10674      0.12090      0.88      0.377
## Year2009         -0.12916      0.11667     -1.11      0.268
## Year2010          0.02039      0.12121      0.17      0.866
## Year2011         -0.00208      0.11704     -0.02      0.986
## Year2012          0.17102      0.12940      1.32      0.186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0303, Adjusted R-squared:  0.0184
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~ = 1. The remaining 1376 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0144 0.8810 0.9420 0.9140 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      6.69e-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 1.078 1          1.038
## Year              1.078 16          1.002

```

Residuals from first author



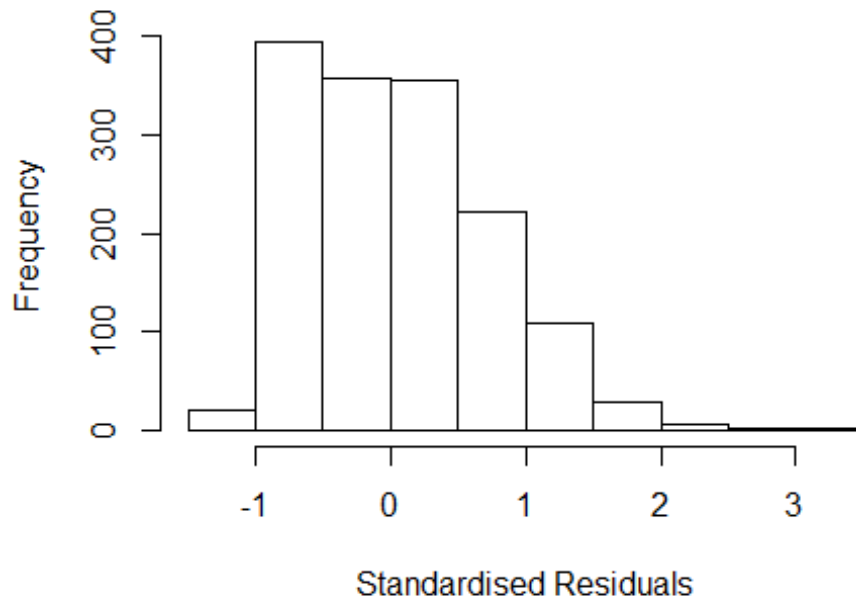
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 169 0031435009 3.522 1997    1710     3     2.935
## 170 0031482189 3.717 1997    1705     4     3.130
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.028 -0.572 -0.024  0.495  3.136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82654    0.09990   8.27 2.9e-16 ***
## FirstAuthorFemale1 -0.00716    0.03809  -0.19  0.851
## Year1997        -0.24592    0.12420  -1.98  0.048 *
## Year1998        -0.17654    0.14439  -1.22  0.222
## Year1999        -0.09555    0.13599  -0.70  0.482
## Year2000        -0.11024    0.14047  -0.78  0.433
## Year2001         0.20123    0.14155   1.42  0.155
## Year2002         0.03102    0.13360   0.23  0.816
## Year2003         0.00255    0.12410   0.02  0.984
## Year2004         0.17596    0.12967   1.36  0.175
## Year2005         0.07762    0.12688   0.61  0.541
```

```

## Year2006          0.11367      0.11775      0.97      0.335
## Year2007          -0.08145      0.11968     -0.68      0.496
## Year2008           0.10337      0.12041      0.86      0.391
## Year2009          -0.13022      0.11640     -1.12      0.263
## Year2010           0.01845      0.12088      0.15      0.879
## Year2011          -0.00533      0.11662     -0.05      0.964
## Year2012           0.16616      0.12885      1.29      0.197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.0188
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1376 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0135 0.8810 0.9420 0.9140 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      6.69e-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 1.058 1      1.028
## Year              1.058 16      1.002

```

Residuals from last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 169 0031435009 3.522 1997    1710      3      2.935
## 170 0031482189 3.717 1997    1705      4      3.130
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0371 -0.5643 -0.0192  0.4946  3.1281
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.30e-01   1.00e-01   8.30 2.3e-16 ***
## LastAuthorFemale1 -2.19e-02   3.76e-02  -0.58  0.560
## Year1997        -2.41e-01   1.24e-01  -1.94  0.053 .
## Year1998        -1.74e-01   1.44e-01  -1.21  0.228
## Year1999        -9.34e-02   1.36e-01  -0.69  0.493
## Year2000        -1.08e-01   1.40e-01  -0.77  0.443
## Year2001         2.07e-01   1.41e-01   1.46  0.144
## Year2002         3.68e-02   1.34e-01   0.28  0.783
## Year2003         7.32e-03   1.24e-01   0.06  0.953
## Year2004         1.79e-01   1.30e-01   1.37  0.170
## Year2005         8.08e-02   1.27e-01   0.63  0.526
```

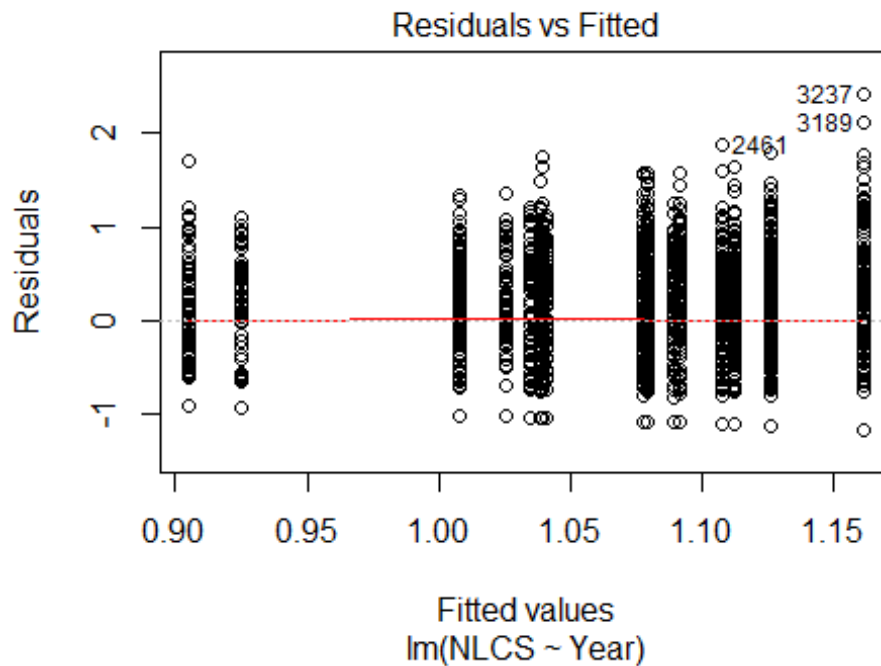


```

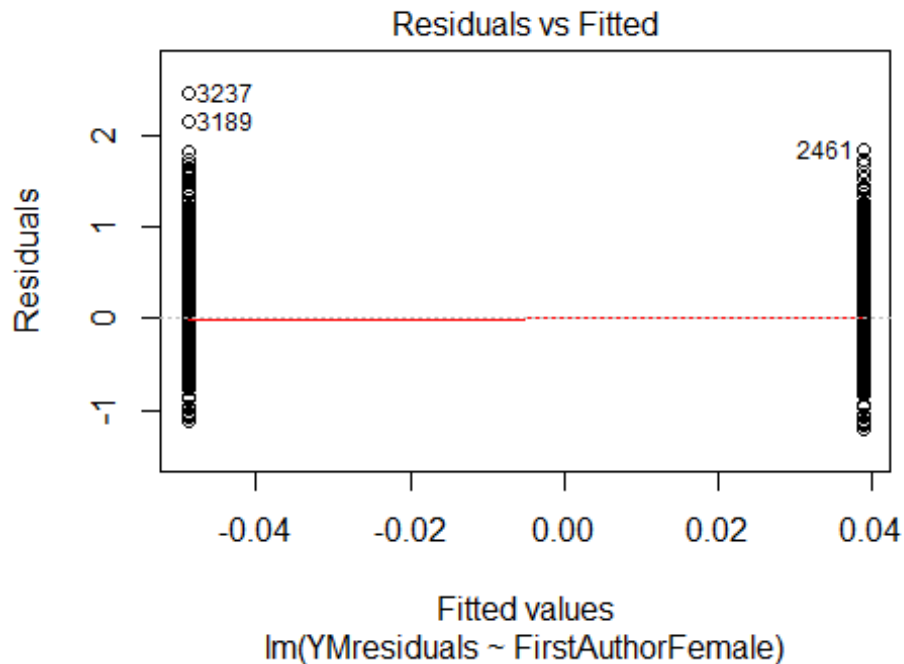
## Year2006      1.16e-01  1.18e-01  0.98  0.326
## Year2007     -7.96e-02  1.20e-01 -0.67  0.506
## Year2008      1.08e-01  1.21e-01  0.89  0.373
## Year2009     -1.27e-01  1.16e-01 -1.09  0.275
## Year2010      2.27e-02  1.21e-01  0.19  0.851
## Year2011     -3.87e-06  1.17e-01  0.00  1.000
## Year2012      1.72e-01  1.29e-01  1.33  0.184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.714
## Multiple R-squared:  0.0301, Adjusted R-squared:  0.0189
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1376 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0159 0.8810 0.9430 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      6.69e-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: 1495"
## [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
## 113 99 101 97 108 124 143 131 156 146 179 192 239 271 235
## 2011 2012
## 327 309
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 99 81 85 85 83 100 118 113 127 126 142 166 192 219 196

```

```
## 2011 2012
## 282 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 77 79 83 81 97 112 110 125 120 136 162 185 210 187
## 2011 2012
## 273 249
## [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
```

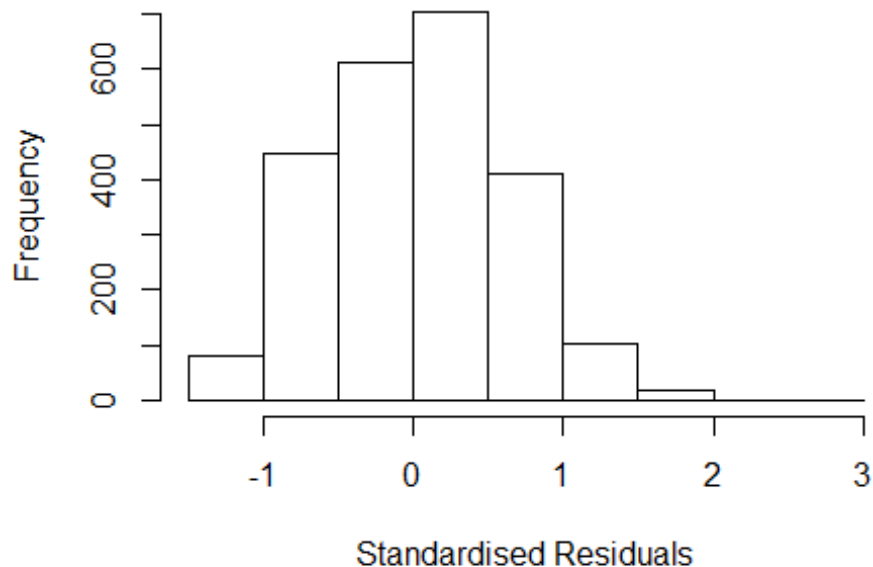


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 1.96021701060058"
## [1] "Male first author team size 2018 geometric mean: 1.53179523289807"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8200, p-value = 0.008
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.86322136866924"
## [1] "Male last author team size 2018 geometric mean: 1.6857436008242"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7700, 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 1.979 1          1.407
## LastAuthorFemale  1.797 1          1.341
## UniqueAuthors    1.370 4          1.040
## Year              1.248 16         1.007
```

Residuals from first and last author and team size



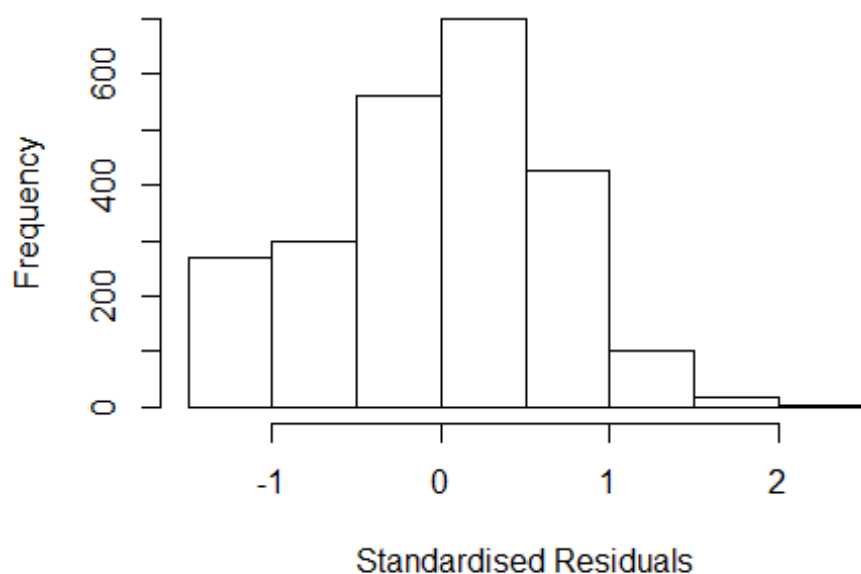
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3237 84859328881 3.575 2012    1203      3    2.602
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3992 -0.4424  0.0329  0.4413  2.6016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8380    0.0602   13.92 < 2e-16 ***
## FirstAuthorFemale1 0.0147    0.0379    0.39  0.699
## LastAuthorFemale1 0.0121    0.0359    0.34  0.736
## UniqueAuthors2    0.2401    0.0357    6.72 2.2e-11 ***
## UniqueAuthors3    0.4026    0.0425    9.48 < 2e-16 ***
## UniqueAuthors4    0.4188    0.0591    7.08 1.9e-12 ***
## UniqueAuthors5    0.5640    0.0522   10.81 < 2e-16 ***
## Year1997          0.0932    0.1009    0.92  0.355
## Year1998         -0.0517    0.0897   -0.58  0.565
## Year1999          0.1029    0.0972    1.06  0.290
```

```

## Year2000          0.0826      0.0875      0.94      0.346
## Year2001          0.1312      0.0871      1.51      0.132
## Year2002          0.0520      0.0887      0.59      0.558
## Year2003          0.1027      0.0964      1.06      0.287
## Year2004          0.1036      0.0814      1.27      0.203
## Year2005          0.1019      0.0806      1.26      0.206
## Year2006          0.0586      0.0821      0.71      0.476
## Year2007          0.1070      0.0780      1.37      0.170
## Year2008          0.0286      0.0750      0.38      0.703
## Year2009          0.1156      0.0704      1.64      0.100
## Year2010          0.0913      0.0737      1.24      0.216
## Year2011          0.1252      0.0715      1.75      0.080 .
## Year2012          0.1354      0.0769      1.76      0.078 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0753, Adjusted R-squared:  0.0667
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 192 weights are ~= 1. The remaining 2188 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0716 0.8440 0.9490 0.9130 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          4.20e-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 2.016 1          1.420
## LastAuthorFemale 1.968 1          1.403
## Year          1.079 16          1.002

```

Residuals from first and last author



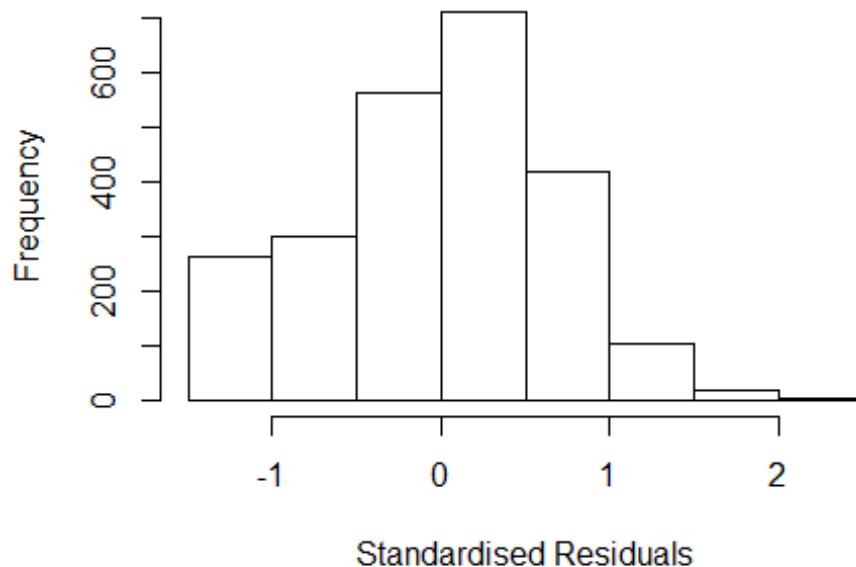
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2101 -0.4598 0.0551 0.4589 2.4927
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8764 0.0625 14.03 < 2e-16 ***
## FirstAuthorFemale1 0.1465 0.0404 3.63 0.00029 ***
## LastAuthorFemale1 -0.0670 0.0394 -1.70 0.08908 .
## Year1997 0.1082 0.1030 1.05 0.29366
## Year1998 -0.0235 0.0940 -0.25 0.80219
## Year1999 0.1148 0.1001 1.15 0.25167
## Year2000 0.1105 0.0950 1.16 0.24496
## Year2001 0.1627 0.0890 1.83 0.06770 .
## Year2002 0.1368 0.0929 1.47 0.14084
## Year2003 0.1546 0.0980 1.58 0.11478
## Year2004 0.1814 0.0835 2.17 0.02983 *
## Year2005 0.1259 0.0839 1.50 0.13334
```

```

## Year2006          0.1051      0.0843      1.25  0.21241
## Year2007          0.1432      0.0820      1.75  0.08098 .
## Year2008          0.0951      0.0779      1.22  0.22240
## Year2009          0.1703      0.0729      2.34  0.01956 *
## Year2010          0.1660      0.0763      2.18  0.02971 *
## Year2011          0.1872      0.0740      2.53  0.01148 *
## Year2012          0.2058      0.0800      2.57  0.01012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.667
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00603
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2180 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.132  0.857   0.948   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      4.20e-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 1.054 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
## -1.1757 -0.4576 0.0472 0.4609 2.4973
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8655 0.0621 13.94 < 2e-16 ***
## FirstAuthorFemale1 0.0981 0.0292 3.36 0.00079 ***
## Year1997 0.1118 0.1031 1.08 0.27831
## Year1998 -0.0191 0.0934 -0.20 0.83835
## Year1999 0.1182 0.1003 1.18 0.23853
## Year2000 0.1096 0.0944 1.16 0.24573
## Year2001 0.1657 0.0892 1.86 0.06344 .
## Year2002 0.1454 0.0927 1.57 0.11689
## Year2003 0.1584 0.0980 1.62 0.10613
## Year2004 0.1854 0.0836 2.22 0.02676 *
## Year2005 0.1324 0.0839 1.58 0.11458
## Year2006 0.1082 0.0843 1.28 0.19905
```

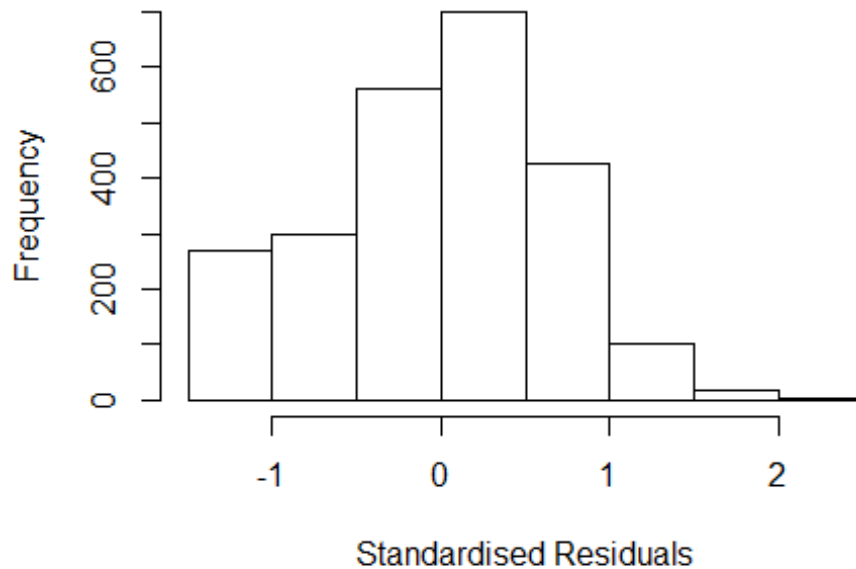


```

## Year2007          0.1463      0.0820      1.79  0.07435 .
## Year2008          0.0987      0.0779      1.27  0.20531
## Year2009          0.1764      0.0727      2.43  0.01535 *
## Year2010          0.1679      0.0764      2.20  0.02807 *
## Year2011          0.1903      0.0739      2.58  0.01008 *
## Year2012          0.2121      0.0798      2.66  0.00792 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.667
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00529
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.130  0.855  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      4.20e-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 1.03 1      1.015
## Year              1.03 16      1.001

```

Residuals from last author



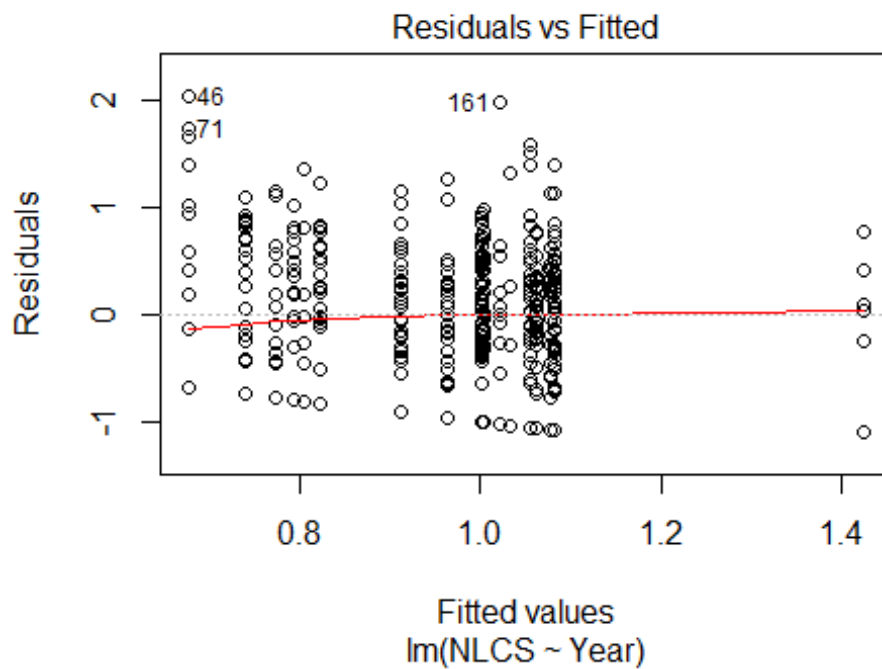
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1562 -0.4647 0.0477 0.4656 2.4556
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8993 0.0624 14.41 <2e-16 ***
## LastAuthorFemale1 0.0367 0.0286 1.28 0.200
## Year1997 0.1126 0.1034 1.09 0.276
## Year1998 -0.0218 0.0935 -0.23 0.816
## Year1999 0.1127 0.1012 1.11 0.265
## Year2000 0.1139 0.0941 1.21 0.226
## Year2001 0.1635 0.0896 1.83 0.068 .
## Year2002 0.1488 0.0932 1.60 0.110
## Year2003 0.1609 0.0984 1.64 0.102
## Year2004 0.1844 0.0839 2.20 0.028 *
## Year2005 0.1320 0.0844 1.57 0.118
## Year2006 0.1086 0.0846 1.28 0.199
```

```

## Year2007          0.1481      0.0823      1.80      0.072 .
## Year2008          0.1023      0.0786      1.30      0.193
## Year2009          0.1865      0.0730      2.56      0.011 *
## Year2010          0.1760      0.0768      2.29      0.022 *
## Year2011          0.1884      0.0742      2.54      0.011 *
## Year2012          0.2201      0.0799      2.75      0.006 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.67
## Multiple R-squared:  0.00797,    Adjusted R-squared:  0.000832
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2171 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.856  0.946  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      4.20e-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: 2380"
## [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
##   22   14   56   41   39   36   38   27   41   31   47   38   39   45   57
## 2011 2012
##   66   78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    6   34   25   17   10   22   24   38   30   39   30   29   34   50
## 2011 2012

```

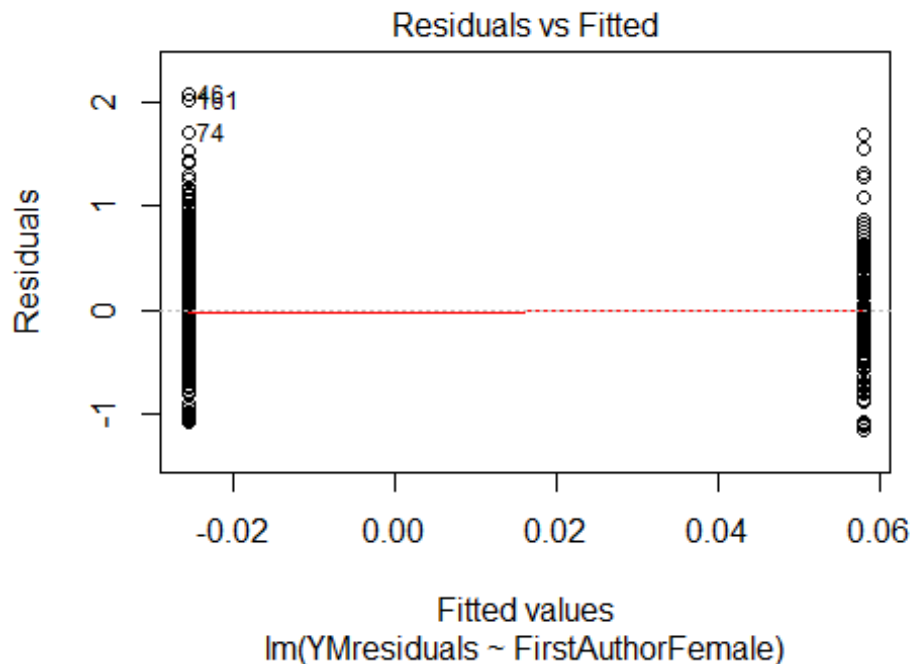
```
## 50 55
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 5 34 23 17 10 21 21 35 29 36 30 27 33 48
## 2011 2012
## 46 54
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.07
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.9, df = 1, p-value = 0.05
## [1] "Female first author team size 2018 geometric mean: 3.15317007128144"
## [1] "Male first author team size 2018 geometric mean: 2.63962895337509"
## 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 = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.55970759249787"
## [1] "Male last author team size 2018 geometric mean: 2.9240639735249"

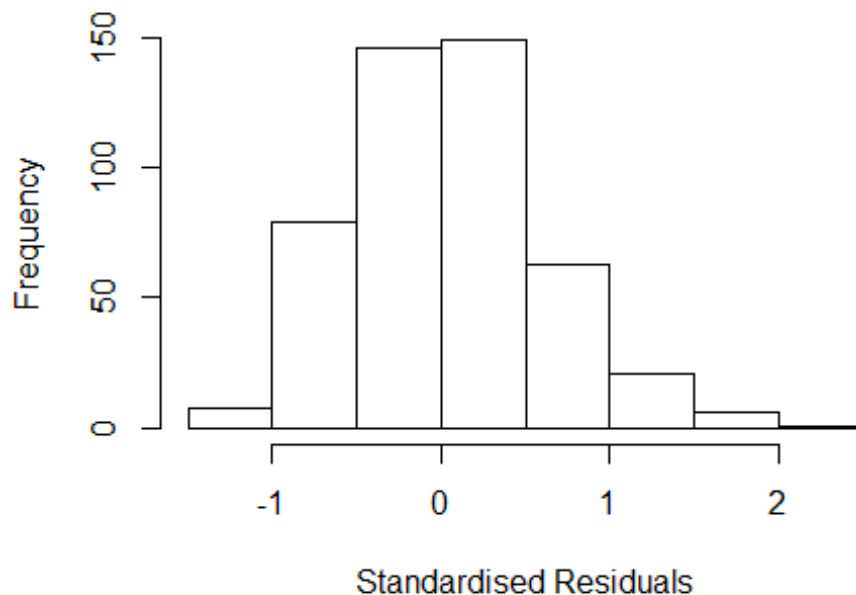
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 0.4
## 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	1.543	1	1.242
LastAuthorFemale	1.442	1	1.201
UniqueAuthors	2.007	4	1.091
Year	2.317	16	1.027

Residuals from first and last author and team size



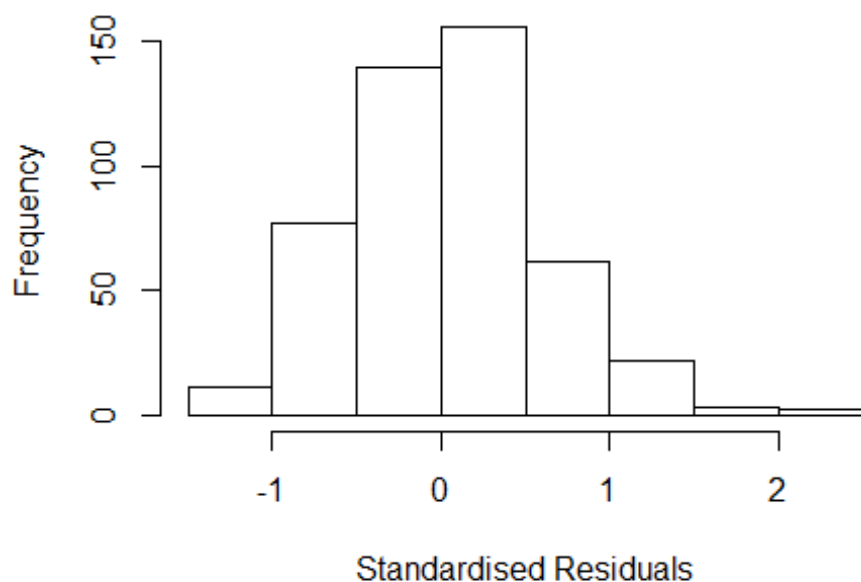
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2993 -0.4011 0.0102 0.3621 2.3169
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8515 0.3130 2.72 0.00677 **
## FirstAuthorFemale1 -0.0268 0.0648 -0.41 0.67925
## LastAuthorFemale1 0.1884 0.0640 2.94 0.00343 **
## UniqueAuthors2 0.3220 0.0677 4.76 2.6e-06 ***
## UniqueAuthors3 0.3126 0.0840 3.72 0.00023 ***
## UniqueAuthors4 0.2999 0.1040 2.88 0.00412 **
## UniqueAuthors5 0.1633 0.1119 1.46 0.14515
## Year1997 0.2354 0.3816 0.62 0.53762
## Year1998 -0.4504 0.3393 -1.33 0.18505
## Year1999 -0.2087 0.3349 -0.62 0.53347
```

```

## Year2000          -0.0752      0.3825    -0.20  0.84428
## Year2001          -0.2631      0.3619    -0.73  0.46751
## Year2002          -0.2010      0.3496    -0.58  0.56550
## Year2003           0.0886      0.3373     0.26  0.79281
## Year2004          -0.3702      0.3240    -1.14  0.25376
## Year2005          -0.1593      0.3268    -0.49  0.62612
## Year2006          -0.1183      0.3215    -0.37  0.71314
## Year2007          -0.2420      0.3306    -0.73  0.46453
## Year2008          -0.0136      0.3261    -0.04  0.96667
## Year2009           0.0354      0.3170     0.11  0.91122
## Year2010           0.0769      0.3187     0.24  0.80942
## Year2011          -0.0773      0.3189    -0.24  0.80858
## Year2012          -0.0532      0.3190    -0.17  0.86775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.171, Adjusted R-squared:  0.13
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 427 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0317 0.8680 0.9470 0.8980 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          2.11e-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 1.507 1          1.227
## LastAuthorFemale  1.419 1          1.191
## Year              1.299 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
## -1.2259 -0.4223 0.0282 0.3929 2.2654
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13240 0.30777 3.68 0.00026 ***
## FirstAuthorFemale1 0.00942 0.06661 0.14 0.88757
## LastAuthorFemale1 0.17893 0.06627 2.70 0.00719 **
## Year1997 0.24352 0.40675 0.60 0.54968
## Year1998 -0.67982 0.33636 -2.02 0.04385 *
## Year1999 -0.35654 0.34222 -1.04 0.29804
## Year2000 -0.26591 0.36755 -0.72 0.46976
## Year2001 -0.43826 0.36941 -1.19 0.23609
## Year2002 -0.38847 0.34689 -1.12 0.26337
## Year2003 -0.15543 0.32987 -0.47 0.63772
## Year2004 -0.55943 0.32755 -1.71 0.08833 .
## Year2005 -0.26599 0.32176 -0.83 0.40886
```

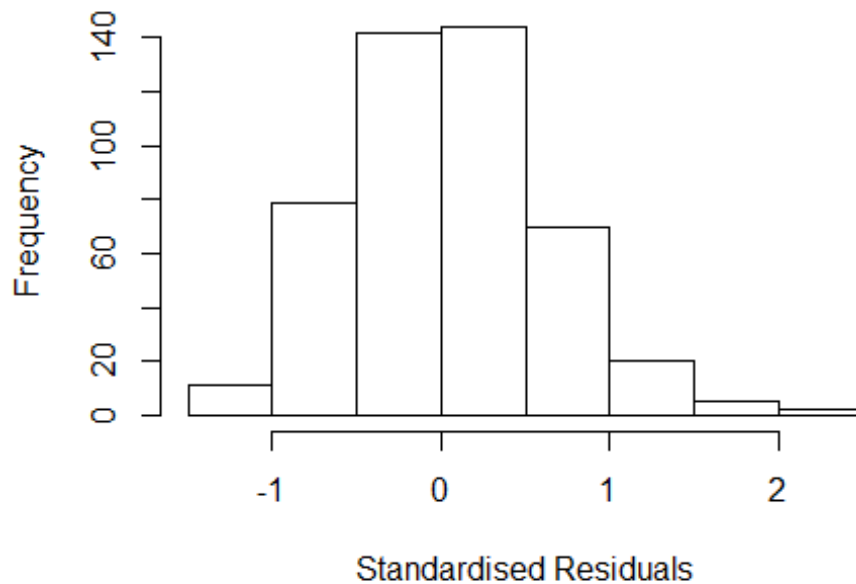


```

## Year2006      -0.27888    0.32143   -0.87  0.38605
## Year2007      -0.36799    0.33640   -1.09  0.27457
## Year2008      -0.17835    0.33031   -0.54  0.58951
## Year2009      -0.09925    0.31955   -0.31  0.75626
## Year2010      -0.09486    0.32295   -0.29  0.76910
## Year2011      -0.20822    0.31926   -0.65  0.51460
## Year2012      -0.16654    0.32277   -0.52  0.60613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0725
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0764 0.8630 0.9500 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      2.11e-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 1.173 1      1.083
## Year      1.173 16      1.005

```

Residuals from first author



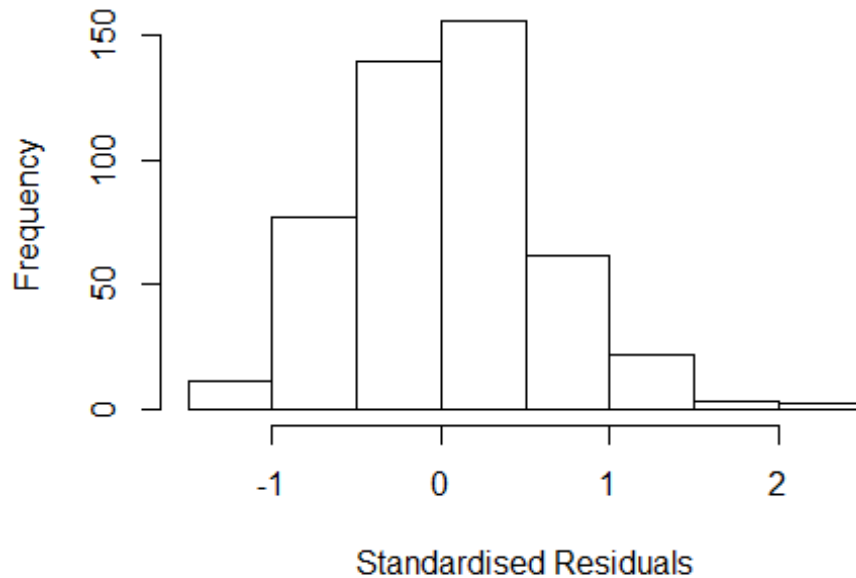
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15188 -0.40028 0.00873 0.38970 2.24670
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2018 0.3384 3.55 0.00042 ***
## FirstAuthorFemale1 0.1028 0.0589 1.75 0.08138 .
## Year1997 0.2474 0.4601 0.54 0.59096
## Year1998 -0.7305 0.3643 -2.00 0.04556 *
## Year1999 -0.4245 0.3689 -1.15 0.25047
## Year2000 -0.2747 0.3917 -0.70 0.48355
## Year2001 -0.4865 0.3956 -1.23 0.21939
## Year2002 -0.4536 0.3728 -1.22 0.22427
## Year2003 -0.1938 0.3591 -0.54 0.58966
## Year2004 -0.6061 0.3565 -1.70 0.08979 .
## Year2005 -0.2910 0.3533 -0.82 0.41059
## Year2006 -0.3275 0.3520 -0.93 0.35266
```

```

## Year2007          -0.4204      0.3675    -1.14   0.25323
## Year2008          -0.2317      0.3613    -0.64   0.52160
## Year2009          -0.1377      0.3498    -0.39   0.69390
## Year2010          -0.1527      0.3530    -0.43   0.66549
## Year2011          -0.2532      0.3500    -0.72   0.46978
## Year2012          -0.2151      0.3534    -0.61   0.54293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.0946, Adjusted R-squared:  0.0608
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0952 0.8700 0.9460 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      2.11e-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 1.132 1          1.064
## Year              1.132 16          1.004

```

Residuals from last author



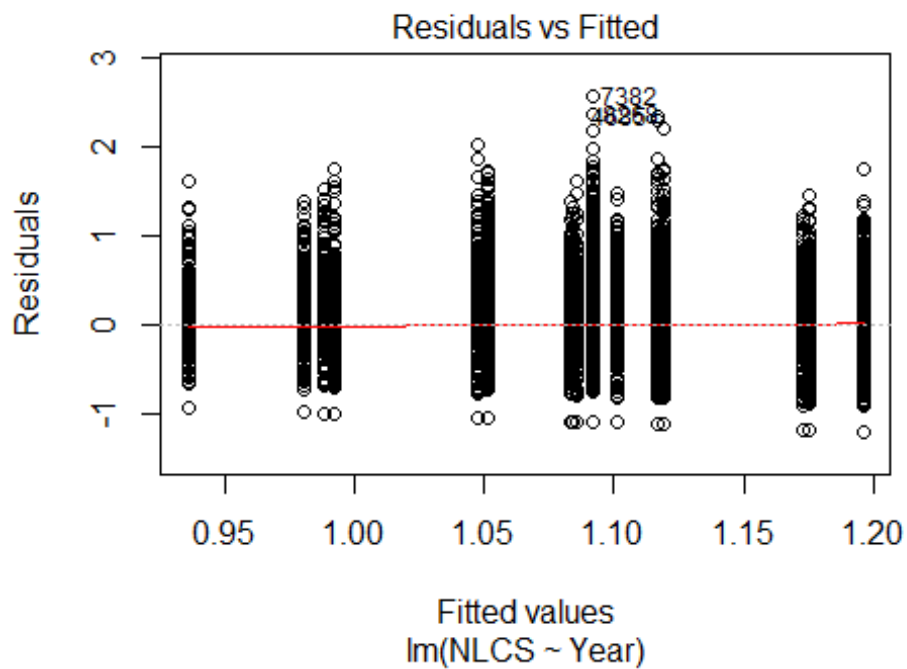
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2240 -0.4229 0.0251 0.3914 2.2660
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1305 0.3071 3.68 0.00026 ***
## LastAuthorFemale1 0.1840 0.0593 3.10 0.00204 **
## Year1997 0.2436 0.4070 0.60 0.54982
## Year1998 -0.6784 0.3351 -2.02 0.04347 *
## Year1999 -0.3544 0.3417 -1.04 0.30021
## Year2000 -0.2643 0.3674 -0.72 0.47237
## Year2001 -0.4357 0.3682 -1.18 0.23726
## Year2002 -0.3846 0.3427 -1.12 0.26223
## Year2003 -0.1520 0.3273 -0.46 0.64261
## Year2004 -0.5557 0.3250 -1.71 0.08799 .
## Year2005 -0.2626 0.3189 -0.82 0.41066
## Year2006 -0.2753 0.3184 -0.86 0.38763
```

```

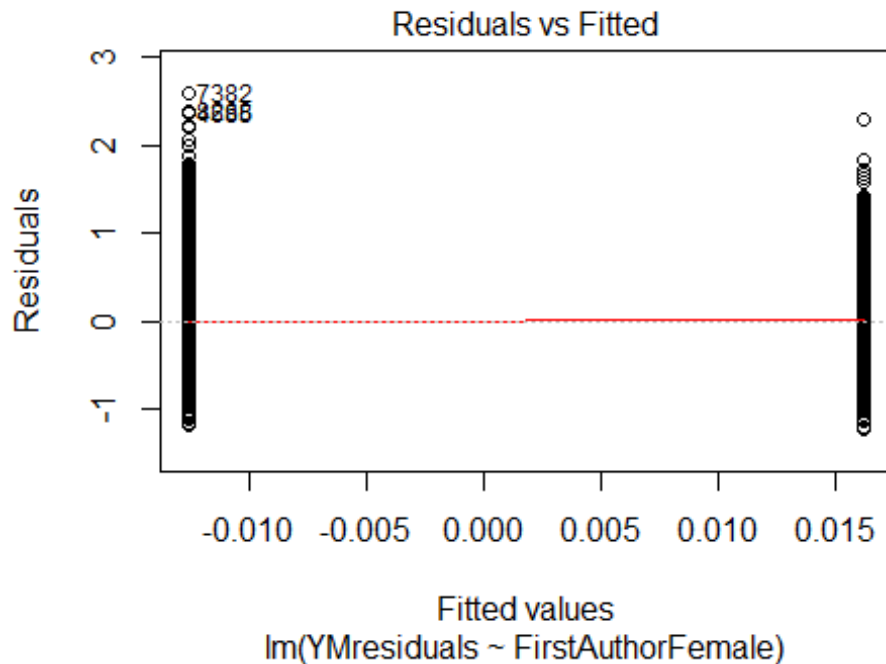
## Year2007          -0.3638      0.3327   -1.09   0.27468
## Year2008          -0.1755      0.3284   -0.53   0.59318
## Year2009          -0.0959      0.3175   -0.30   0.76282
## Year2010          -0.0905      0.3196   -0.28   0.77710
## Year2011          -0.2052      0.3174   -0.65   0.51821
## Year2012          -0.1619      0.3183   -0.51   0.61121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0748
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 442 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0739 0.8620 0.9490 0.9030 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      2.11e-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: 473"
## [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
## 225 197 268 225 310 281 330 267 295 361 381 465 734 824 924
## 2011 2012
## 926 885
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 191 166 216 195 247 226 276 231 251 316 323 418 636 701 806
## 2011 2012

```

```
## 819 731
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 186 158 211 191 240 217 266 222 241 304 306 405 610 680 757
## 2011 2012
## 787 706
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 1e-04
```

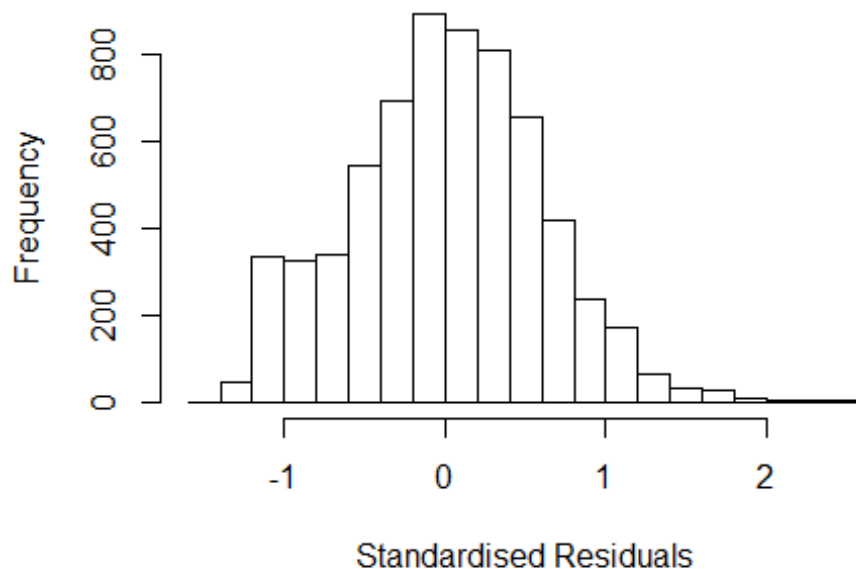


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 1.83716037382447"
## [1] "Male first author team size 2018 geometric mean: 1.42476143632355"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140000, p-value = 3e-11
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.70241824898697"
## [1] "Male last author team size 2018 geometric mean: 1.58417544221344"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120000, 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 2.081 1 1.443
## LastAuthorFemale 2.061 1 1.436
## UniqueAuthors 1.096 4 1.012
## Year 1.090 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
## -1.4050 -0.3916 0.0138 0.4021 2.4760
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.16e-01 4.58e-02 19.99 < 2e-16 ***
## FirstAuthorFemale1 -9.24e-03 2.20e-02 -0.42 0.67423
## LastAuthorFemale1 3.40e-02 2.20e-02 1.55 0.12211
## UniqueAuthors2 1.70e-01 1.96e-02 8.67 < 2e-16 ***
## UniqueAuthors3 2.68e-01 2.68e-02 9.98 < 2e-16 ***
## UniqueAuthors4 3.46e-01 4.43e-02 7.81 6.5e-15 ***
## UniqueAuthors5 2.16e-01 5.27e-02 4.10 4.2e-05 ***
## Year1997 -4.89e-02 6.55e-02 -0.75 0.45584
## Year1998 2.34e-03 6.22e-02 0.04 0.97002
## Year1999 1.18e-01 5.97e-02 1.97 0.04876 *
```

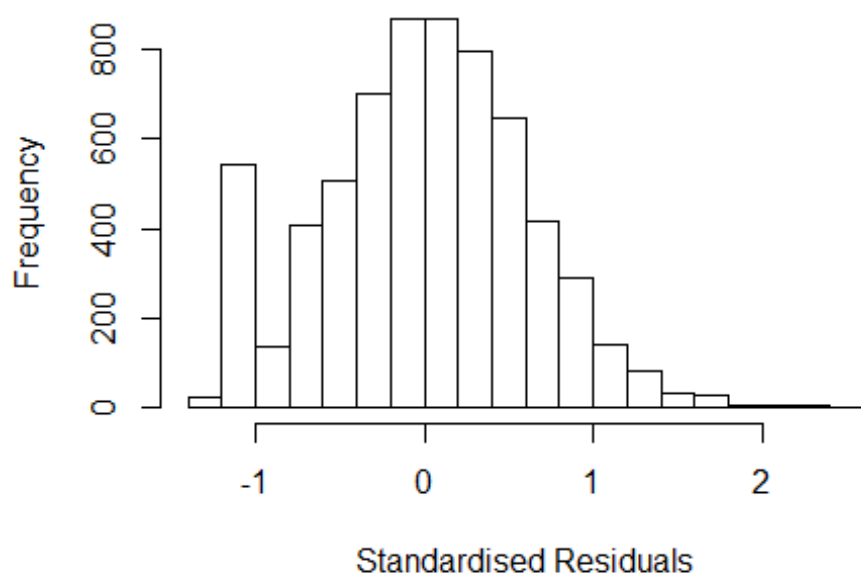


```

## Year2000          3.84e-05    6.01e-02    0.00  0.99949
## Year2001          1.02e-01    5.95e-02    1.71  0.08699 .
## Year2002          1.11e-01    5.87e-02    1.90  0.05789 .
## Year2003          1.07e-01    6.12e-02    1.74  0.08104 .
## Year2004          1.78e-01    5.71e-02    3.13  0.00177 **
## Year2005          4.53e-02    5.76e-02    0.79  0.43124
## Year2006          2.11e-01    5.70e-02    3.70  0.00022 ***
## Year2007          1.94e-01    5.39e-02    3.60  0.00032 ***
## Year2008          1.12e-01    5.13e-02    2.19  0.02872 *
## Year2009          1.06e-01    5.10e-02    2.07  0.03825 *
## Year2010          3.85e-02    5.07e-02    0.76  0.44730
## Year2011          6.89e-02    5.15e-02    1.34  0.18120
## Year2012          6.58e-02    5.26e-02    1.25  0.21090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0403, Adjusted R-squared:  0.037
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 505 weights are ~= 1. The remaining 5982 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0368 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          1.54e-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 2.145 1          1.465
## LastAuthorFemale 2.145 1          1.465
## Year          1.020 16          1.001

```

Residuals from first and last author



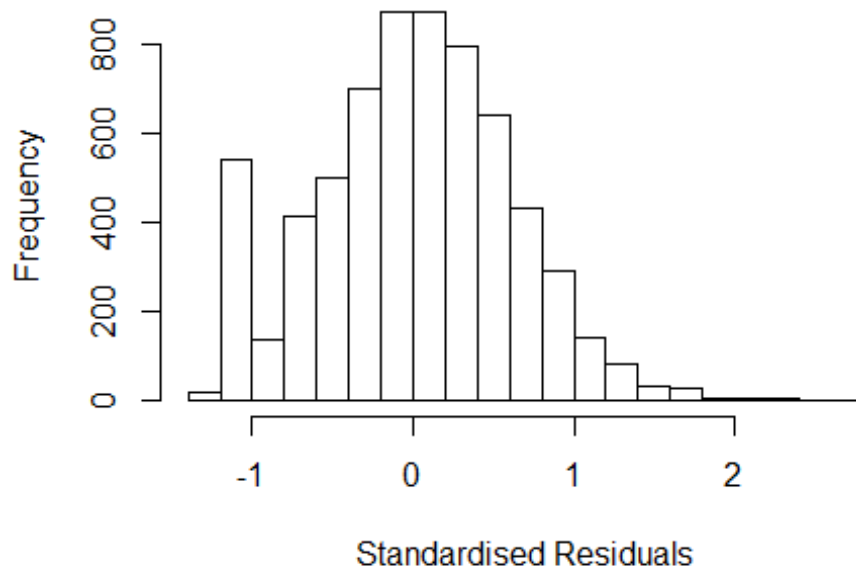
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7382 79955879153 3.665 2011      3312      2      2.584
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2247 -0.3934  0.0148  0.4060  2.5842
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9438    0.0461   20.48 < 2e-16 ***
## FirstAuthorFemale1 0.0210    0.0227    0.93  0.35476
## LastAuthorFemale1 0.0256    0.0229    1.12  0.26265
## Year1997        -0.0434    0.0654   -0.66  0.50678
## Year1998         0.0099    0.0622    0.16  0.87361
## Year1999         0.1247    0.0602    2.07  0.03814 *
## Year2000         0.0158    0.0609    0.26  0.79584
## Year2001         0.1208    0.0604    2.00  0.04528 *
## Year2002         0.1327    0.0593    2.24  0.02535 *
## Year2003         0.1256    0.0614    2.05  0.04073 *
## Year2004         0.2094    0.0573    3.66  0.00026 ***
## Year2005         0.0668    0.0580    1.15  0.24958
```

```

## Year2006          0.2342      0.0573      4.09  4.4e-05 ***
## Year2007          0.2294      0.0542      4.24  2.3e-05 ***
## Year2008          0.1399      0.0516      2.71  0.00676 **
## Year2009          0.1394      0.0514      2.71  0.00670 **
## Year2010          0.0700      0.0512      1.37  0.17116
## Year2011          0.1114      0.0518      2.15  0.03164 *
## Year2012          0.1135      0.0530      2.14  0.03230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0125, Adjusted R-squared:  0.0098
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 536 weights are ~= 1. The remaining 5951 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0229 0.8690 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      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 1.013 1      1.007
## Year              1.013 16      1.000

```

Residuals from first author



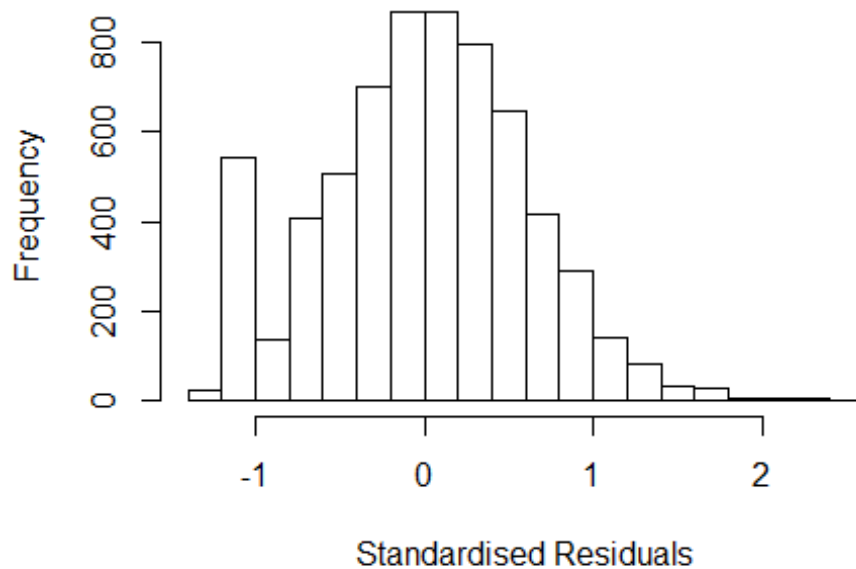
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7382 79955879153 3.665 2011      3312      2      2.584
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2200 -0.3942  0.0141  0.4065  2.6073
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9464    0.0460   20.58 < 2e-16 ***
## FirstAuthorFemale1 0.0399    0.0156    2.55  0.01081 *
## Year1997      -0.0442    0.0654   -0.68  0.49908
## Year1998       0.0095    0.0622    0.15  0.87862
## Year1999       0.1240    0.0601    2.06  0.03918 *
## Year2000       0.0154    0.0609    0.25  0.80094
## Year2001       0.1200    0.0603    1.99  0.04681 *
## Year2002       0.1325    0.0593    2.23  0.02554 *
## Year2003       0.1252    0.0614    2.04  0.04124 *
## Year2004       0.2094    0.0572    3.66  0.00026 ***
## Year2005       0.0670    0.0580    1.15  0.24852
## Year2006       0.2337    0.0573    4.08  4.6e-05 ***
```

```

## Year2007          0.2301      0.0541      4.25  2.2e-05 ***
## Year2008          0.1401      0.0516      2.71  0.00666 **
## Year2009          0.1392      0.0513      2.71  0.00673 **
## Year2010          0.0704      0.0511      1.38  0.16869
## Year2011          0.1113      0.0518      2.15  0.03178 *
## Year2012          0.1135      0.0530      2.14  0.03210 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0123, Adjusted R-squared:  0.00974
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 537 weights are ~= 1. The remaining 5950 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0189 0.8690 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      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 1.013 1      1.006
## Year              1.013 16      1.000

```

Residuals from last author



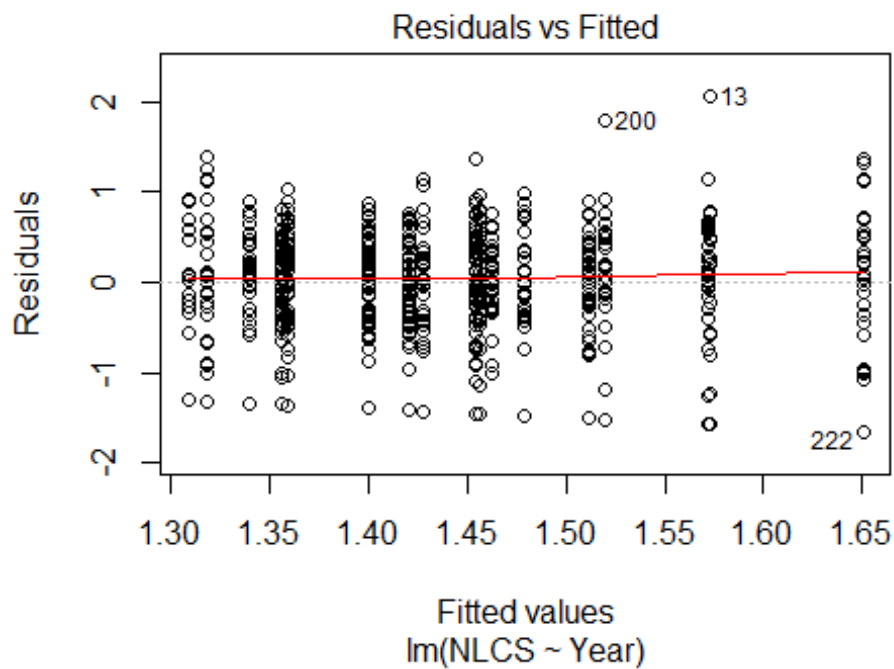
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7382 79955879153 3.665 2011      3312      2      2.584
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2219 -0.3944  0.0155  0.4075  2.5656
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9453     0.0460  20.54 < 2e-16 ***
## LastAuthorFemale1  0.0412     0.0157   2.62  0.00872 **
## Year1997        -0.0431     0.0654  -0.66  0.50978
## Year1998         0.0105     0.0622   0.17  0.86603
## Year1999         0.1257     0.0601   2.09  0.03666 *
## Year2000         0.0167     0.0609   0.27  0.78403
## Year2001         0.1219     0.0604   2.02  0.04337 *
## Year2002         0.1333     0.0593   2.25  0.02472 *
## Year2003         0.1266     0.0614   2.06  0.03928 *
## Year2004         0.2105     0.0573   3.67  0.00024 ***
## Year2005         0.0681     0.0580   1.17  0.24095
## Year2006         0.2354     0.0573   4.11  4.0e-05 ***
```

```

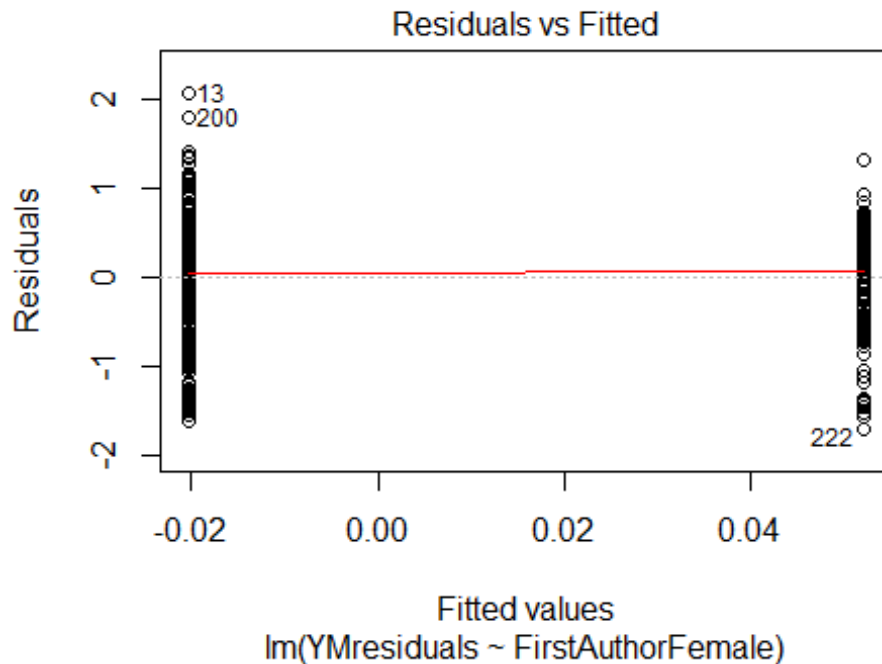
## Year2007          0.2306      0.0542      4.26  2.1e-05 ***
## Year2008          0.1409      0.0516      2.73  0.00639 **
## Year2009          0.1411      0.0514      2.75  0.00605 **
## Year2010          0.0710      0.0512      1.39  0.16564
## Year2011          0.1128      0.0518      2.18  0.02957 *
## Year2012          0.1148      0.0530      2.17  0.03041 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00982
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 529 weights are ~= 1. The remaining 5958 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0266 0.8690 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      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: 6487"
## [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
##   21   24   37   37   32   45   35   49   35   32   59   58   71   48   83
## 2011 2012
##   74   85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   17   27   28   28   30   28   31   28   27   41   47   57   37   60
## 2011 2012

```

```
## 53 60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 16 27 24 24 26 25 31 23 25 40 43 51 33 54
## 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 = 41, df = 16, p-value = 6e-04
```

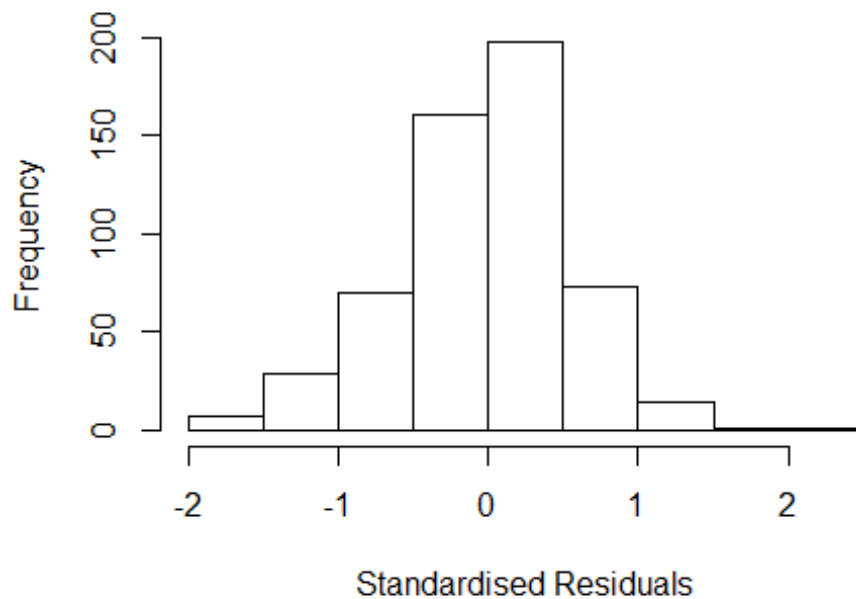


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.4, df = 1, p-value = 0.06
```

```
## [1] "Female first author team size 2018 geometric mean: 3.02180137255297"
## [1] "Male first author team size 2018 geometric mean: 2.50658804162255"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.74439451677223"
## [1] "Male last author team size 2018 geometric mean: 2.63976722382644"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, 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 1.240 1          1.113
## LastAuthorFemale  1.234 1          1.111
## UniqueAuthors     1.952 4          1.087
## Year              2.164 16         1.024
```

Residuals from first and last author and team size



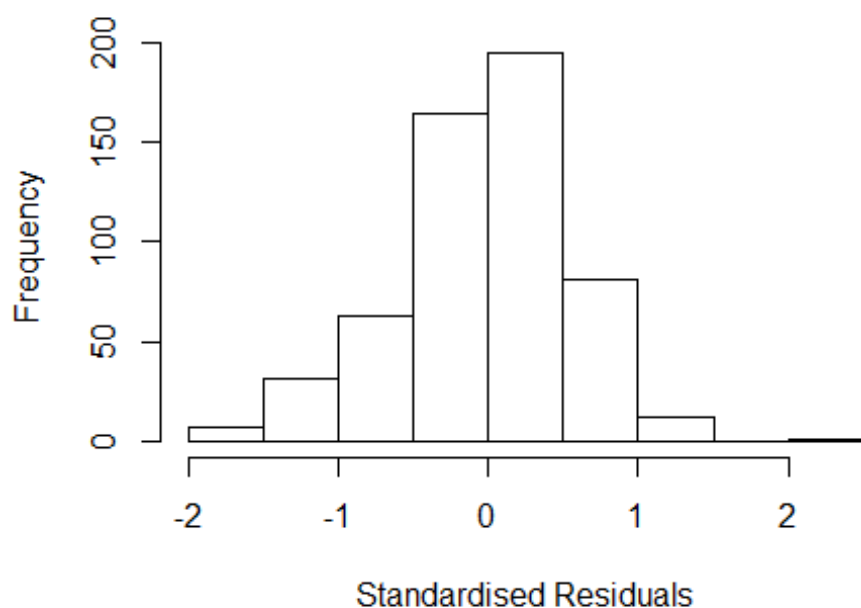
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7086 -0.3379 0.0176 0.3390 2.1635
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32162 0.27708 4.77 2.4e-06 ***
## FirstAuthorFemale1 0.05406 0.05244 1.03 0.303
## LastAuthorFemale1 0.04819 0.05605 0.86 0.390
## UniqueAuthors2 0.14188 0.06221 2.28 0.023 *
## UniqueAuthors3 0.17074 0.07091 2.41 0.016 *
## UniqueAuthors4 0.12711 0.09836 1.29 0.197
## UniqueAuthors5 0.09946 0.10039 0.99 0.322
## Year1997 -0.06385 0.31489 -0.20 0.839
## Year1998 -0.12056 0.34599 -0.35 0.728
## Year1999 0.14454 0.30312 0.48 0.634
```

```

## Year2000      0.24441    0.29700    0.82    0.411
## Year2001      0.05564    0.30441    0.18    0.855
## Year2002      0.28474    0.35482    0.80    0.423
## Year2003     -0.00319    0.29292   -0.01    0.991
## Year2004      0.09995    0.29096    0.34    0.731
## Year2005     -0.01727    0.29842   -0.06    0.954
## Year2006      0.03033    0.29135    0.10    0.917
## Year2007     -0.05170    0.28949   -0.18    0.858
## Year2008     -0.08820    0.28744   -0.31    0.759
## Year2009      0.01682    0.29168    0.06    0.954
## Year2010     -0.04952    0.29170   -0.17    0.865
## Year2011     -0.00785    0.29234   -0.03    0.979
## Year2012     -0.00468    0.29029   -0.02    0.987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.523
## Multiple R-squared:  0.05,   Adjusted R-squared:  0.0107
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0485 0.8670 0.9550 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      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 1.193 1      1.092
## LastAuthorFemale 1.198 1      1.094
## Year      1.205 16      1.006

```

Residuals from first and last author

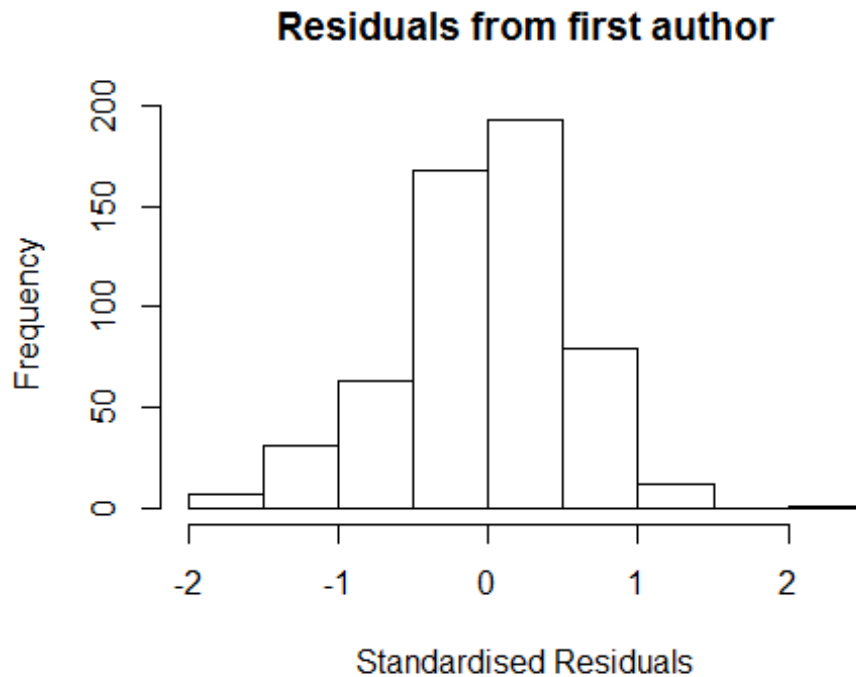


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8001 -0.3592 0.0308 0.3418 2.2450
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.38205 0.29866 4.63 4.6e-06 ***
## FirstAuthorFemale1 0.07226 0.05272 1.37 0.17
## LastAuthorFemale1 0.04867 0.05648 0.86 0.39
## Year1997 -0.06551 0.33445 -0.20 0.84
## Year1998 -0.10661 0.35606 -0.30 0.76
## Year1999 0.14277 0.32218 0.44 0.66
## Year2000 0.28075 0.31470 0.89 0.37
## Year2001 0.07273 0.32178 0.23 0.82
## Year2002 0.29709 0.37055 0.80 0.42
## Year2003 -0.00535 0.31381 -0.02 0.99
## Year2004 0.09546 0.31223 0.31 0.76
## Year2005 0.01107 0.31881 0.03 0.97
```

```

## Year2006          0.04664      0.31008      0.15      0.88
## Year2007          -0.00821      0.30745     -0.03      0.98
## Year2008          -0.07093      0.30726     -0.23      0.82
## Year2009           0.03663      0.31264      0.12      0.91
## Year2010          -0.00559      0.30978     -0.02      0.99
## Year2011           0.05467      0.30942      0.18      0.86
## Year2012           0.02165      0.30832      0.07      0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.0354, Adjusted R-squared:  0.00298
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0306 0.8730 0.9530 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      1.81e-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 1.09 1      1.044
## Year              1.09 16      1.003

```

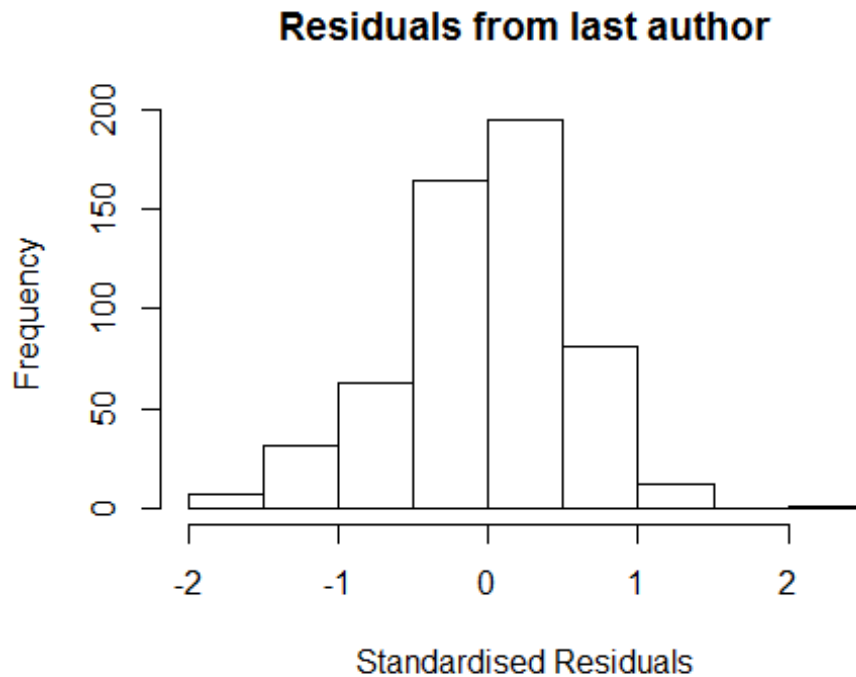


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7846 -0.3572  0.0399  0.3450  2.2431
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.38387    0.30141    4.59 5.5e-06 ***
## FirstAuthorFemale1 0.09111    0.05069    1.80  0.073 .
## Year1997      -0.05623    0.33647   -0.17  0.867
## Year1998      -0.10653    0.35788   -0.30  0.766
## Year1999       0.14586    0.32464    0.45  0.653
## Year2000       0.28392    0.31761    0.89  0.372
## Year2001       0.07200    0.32523    0.22  0.825
## Year2002       0.30963    0.37186    0.83  0.405
## Year2003      -0.00288    0.31621   -0.01  0.993
## Year2004       0.09616    0.31482    0.31  0.760
## Year2005       0.01820    0.32095    0.06  0.955
## Year2006       0.05610    0.31213    0.18  0.857
```

```

## Year2007          -0.00387    0.31011   -0.01    0.990
## Year2008          -0.06449    0.30985   -0.21    0.835
## Year2009           0.04270    0.31513    0.14    0.892
## Year2010          -0.00212    0.31225   -0.01    0.995
## Year2011           0.06198    0.31159    0.20    0.842
## Year2012           0.02505    0.31105    0.08    0.936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0342, Adjusted R-squared:  0.00353
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 518 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0328 0.8760 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      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 1.096 1          1.047
## Year              1.096 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
## -1.764 -0.344 0.026 0.332 2.243
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.384388 0.302731 4.57 6e-06 ***
## LastAuthorFemale1 0.079356 0.054063 1.47 0.14
## Year1997 -0.069625 0.338198 -0.21 0.84
## Year1998 -0.094286 0.359135 -0.26 0.79
## Year1999 0.154553 0.327143 0.47 0.64
## Year2000 0.288898 0.317650 0.91 0.36
## Year2001 0.090008 0.323896 0.28 0.78
## Year2002 0.300489 0.374739 0.80 0.42
## Year2003 0.004616 0.317441 0.01 0.99
## Year2004 0.111687 0.315348 0.35 0.72
## Year2005 0.020609 0.321649 0.06 0.95
## Year2006 0.061247 0.313729 0.20 0.85
```

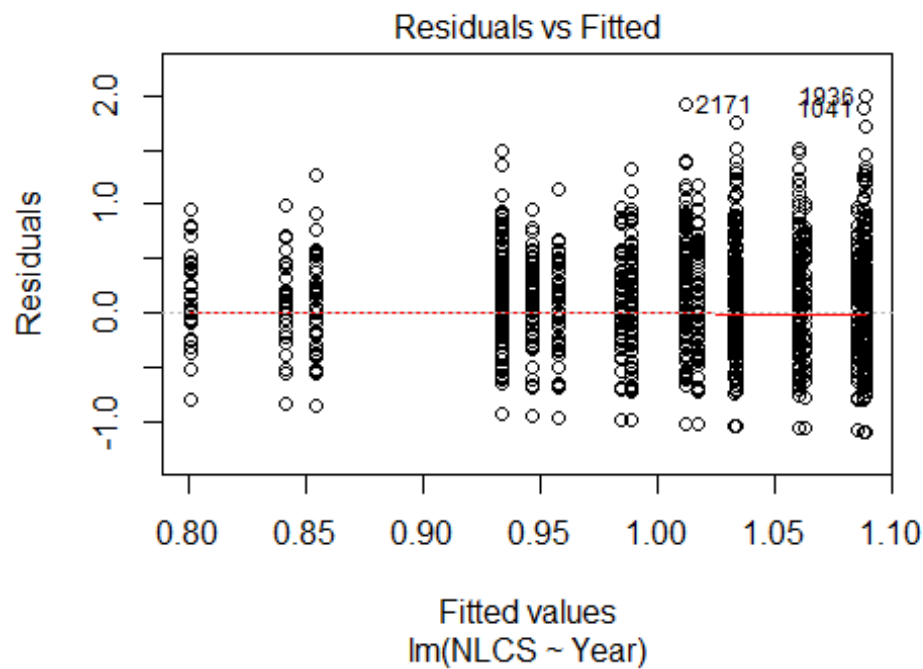


```

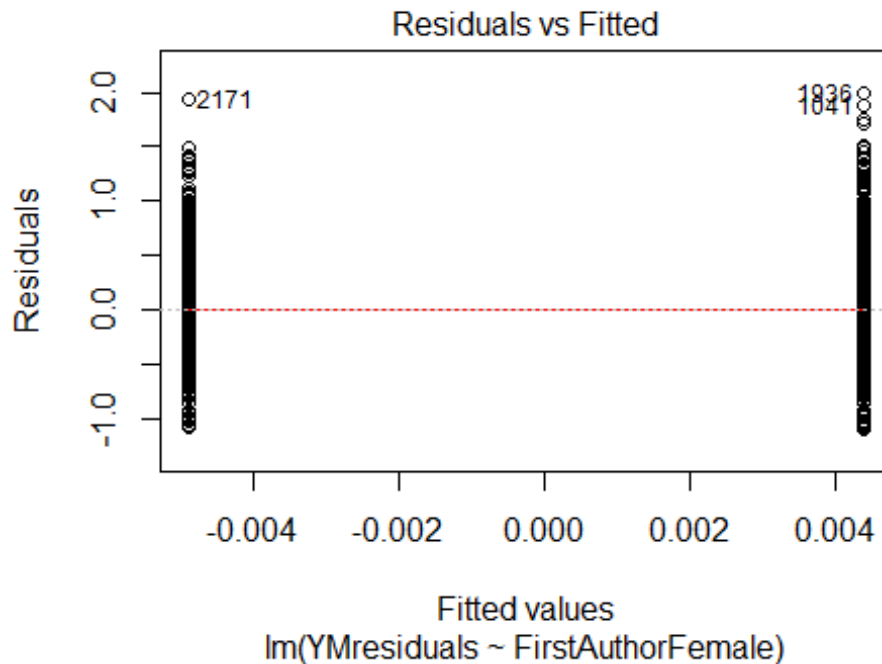
## Year2007      0.000833  0.311290  0.00  1.00
## Year2008     -0.062598  0.310888 -0.20  0.84
## Year2009      0.055596  0.315461  0.18  0.86
## Year2010      0.014031  0.313007  0.04  0.96
## Year2011      0.065164  0.313502  0.21  0.84
## Year2012      0.034192  0.312180  0.11  0.91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.0327, Adjusted R-squared:  0.00198
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.030  0.873  0.952  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      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: 554"
## [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
##   65   51   60   68   91   82   61   98   96   96  124  138  160  185  175
## 2011 2012
##  177  206
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   53   37   42   57   72   64   55   86   65   79   90  112  127  155  147
## 2011 2012

```

```
## 149 174
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 37 42 56 70 59 53 83 64 78 88 107 123 151 141
## 2011 2012
## 143 172
## [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
```

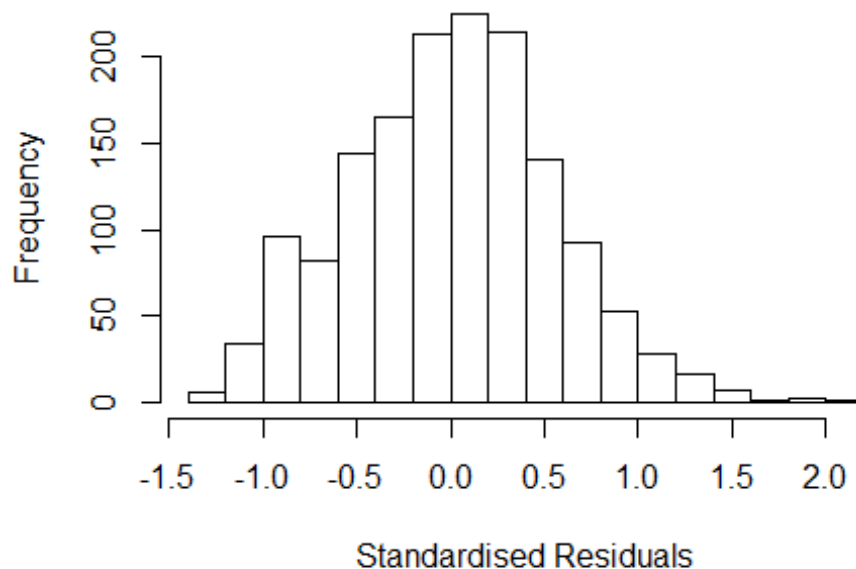


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 1.47159937517692"
## [1] "Male first author team size 2018 geometric mean: 1.32093294512468"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6400, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.39171190344638"
## [1] "Male last author team size 2018 geometric mean: 1.42648302695129"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5700, 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 2.438 1      1.562
## LastAuthorFemale  2.444 1      1.563
## UniqueAuthors    1.238 4      1.027
## Year             1.317 16      1.009
```

Residuals from first and last author and team size



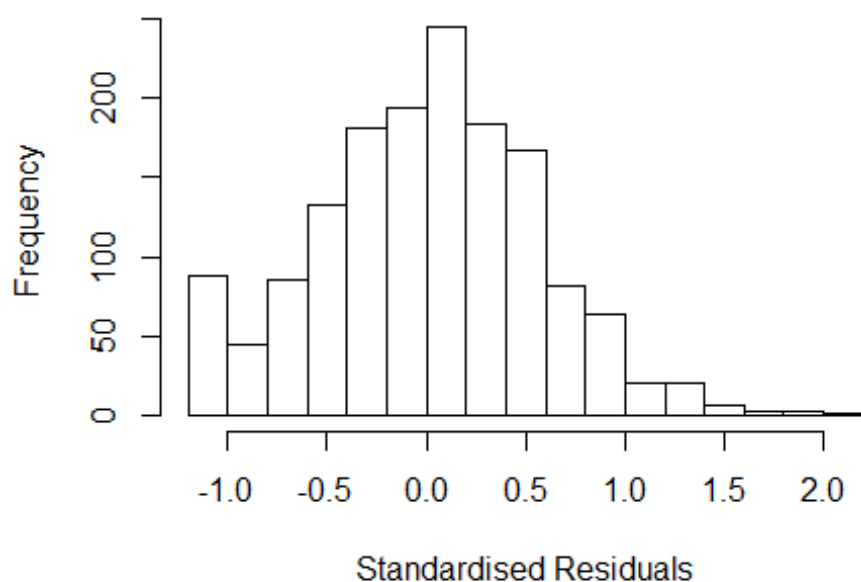
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2365 -0.3678 0.0173 0.3554 2.0791
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82420 0.06690 12.32 < 2e-16 ***
## FirstAuthorFemale1 -0.00815 0.04476 -0.18 0.856
## LastAuthorFemale1 0.00404 0.04486 0.09 0.928
## UniqueAuthors2 0.19361 0.03978 4.87 1.3e-06 ***
## UniqueAuthors3 0.28712 0.05974 4.81 1.7e-06 ***
## UniqueAuthors4 0.42569 0.08084 5.27 1.6e-07 ***
## UniqueAuthors5 0.09423 0.15139 0.62 0.534
## Year1997 -0.02134 0.09791 -0.22 0.827
## Year1998 -0.08915 0.10188 -0.88 0.382
## Year1999 0.10132 0.09173 1.10 0.270
```

```

## Year2000      0.06675    0.08875    0.75    0.452
## Year2001      0.08881    0.08511    1.04    0.297
## Year2002      0.21310    0.10386    2.05    0.040 *
## Year2003      0.11225    0.08671    1.29    0.196
## Year2004      0.16863    0.09168    1.84    0.066 .
## Year2005      0.11869    0.09595    1.24    0.216
## Year2006      0.20846    0.08675    2.40    0.016 *
## Year2007      0.16581    0.08306    2.00    0.046 *
## Year2008      0.04275    0.08480    0.50    0.614
## Year2009      0.17783    0.08342    2.13    0.033 *
## Year2010      0.12113    0.08484    1.43    0.154
## Year2011      0.17872    0.09085    1.97    0.049 *
## Year2012      0.12142    0.08375    1.45    0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0517, Adjusted R-squared:  0.0378
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1400 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.114  0.872  0.951  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      6.58e-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 2.750 1      1.658
## LastAuthorFemale 2.763 1      1.662
## Year      1.076 16      1.002

```

Residuals from first and last author



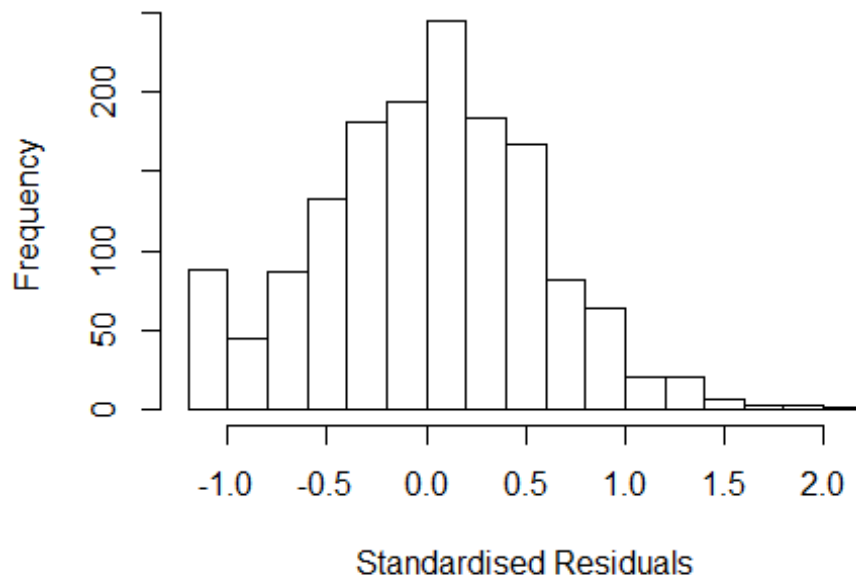
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0875 -0.3684 0.0211 0.3800 2.0216
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.846875 0.067408 12.56 <2e-16 ***
## FirstAuthorFemale1 0.004680 0.048593 0.10 0.9233
## LastAuthorFemale1 -0.000393 0.048689 -0.01 0.9936
## Year1997 -0.008167 0.099529 -0.08 0.9346
## Year1998 -0.043447 0.101535 -0.43 0.6688
## Year1999 0.116413 0.092918 1.25 0.2105
## Year2000 0.104316 0.090051 1.16 0.2469
## Year2001 0.108037 0.088501 1.22 0.2224
## Year2002 0.220863 0.104512 2.11 0.0347 *
## Year2003 0.176106 0.086444 2.04 0.0418 *
## Year2004 0.186494 0.095190 1.96 0.0503 .
## Year2005 0.128788 0.094766 1.36 0.1743
```

```

## Year2006          0.235940    0.088033    2.68    0.0074 **
## Year2007          0.199123    0.085559    2.33    0.0201 *
## Year2008          0.078916    0.085406    0.92    0.3556
## Year2009          0.221737    0.083920    2.64    0.0083 **
## Year2010          0.160554    0.084292    1.90    0.0570 .
## Year2011          0.213510    0.092813    2.30    0.0216 *
## Year2012          0.160462    0.084212    1.91    0.0569 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.00358
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.165  0.873   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      6.58e-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 1.05 1         1.024
## Year              1.05 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
## -1.0871 -0.3682 0.0211 0.3800 2.0216
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.84686 0.06737 12.57 <2e-16 ***
## FirstAuthorFemale1 0.00437 0.03002 0.15 0.8844
## Year1997 -0.00818 0.09950 -0.08 0.9345
## Year1998 -0.04345 0.10153 -0.43 0.6687
## Year1999 0.11640 0.09291 1.25 0.2105
## Year2000 0.10427 0.09006 1.16 0.2471
## Year2001 0.10802 0.08849 1.22 0.2224
## Year2002 0.22085 0.10450 2.11 0.0347 *
## Year2003 0.17609 0.08650 2.04 0.0420 *
## Year2004 0.18648 0.09517 1.96 0.0503 .
## Year2005 0.12878 0.09477 1.36 0.1744
## Year2006 0.23591 0.08794 2.68 0.0074 **
```

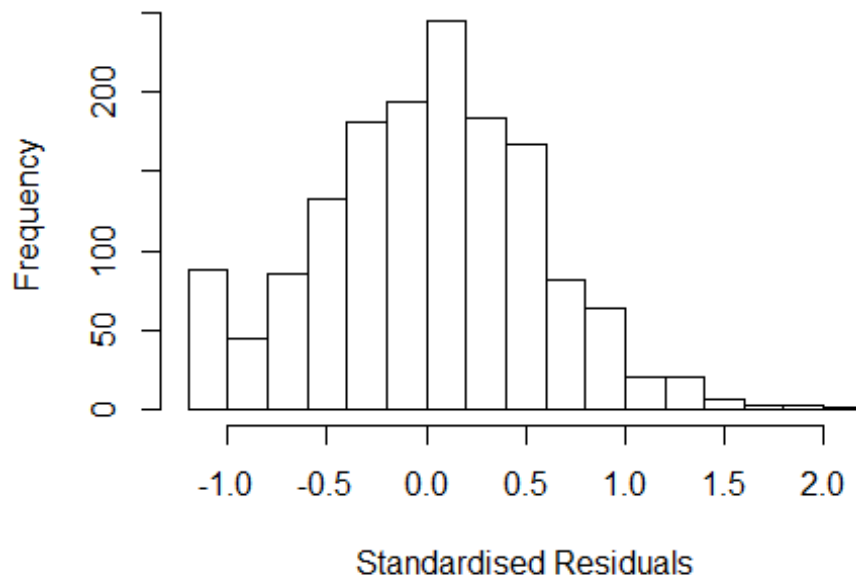


```

## Year2007          0.19912      0.08552      2.33      0.0200 *
## Year2008          0.07889      0.08533      0.92      0.3554
## Year2009          0.22174      0.08391      2.64      0.0083 **
## Year2010          0.16053      0.08419      1.91      0.0568 .
## Year2011          0.21351      0.09277      2.30      0.0215 *
## Year2012          0.16043      0.08404      1.91      0.0565 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.00424
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.165  0.873  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      6.58e-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 1.055 1          1.027
## Year            1.055 16          1.002

```

Residuals from last author



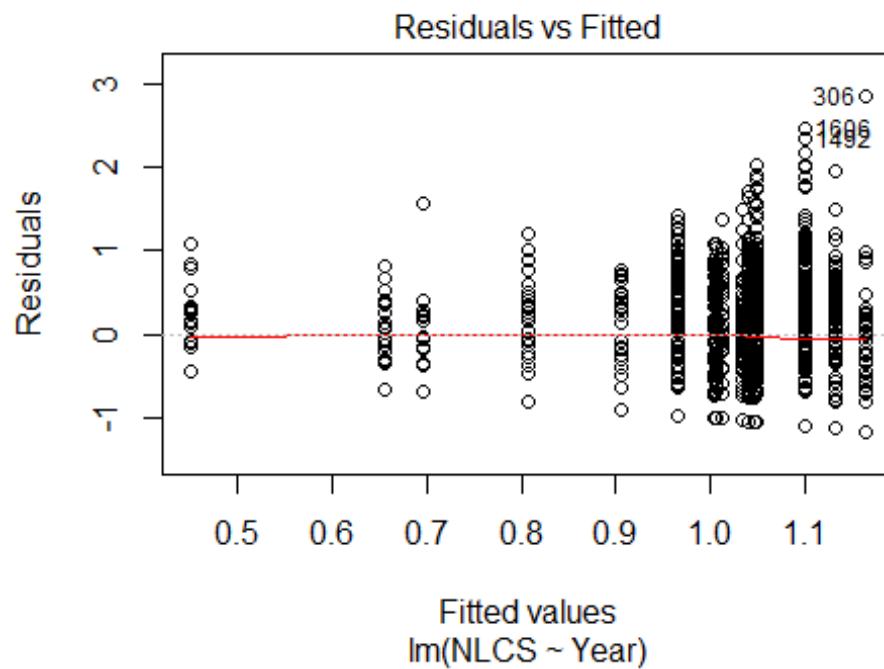
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0865 -0.3682 0.0211 0.3799 2.0211
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.84727 0.06734 12.58 <2e-16 ***
## LastAuthorFemale1 0.00329 0.03008 0.11 0.9129
## Year1997 -0.00817 0.09955 -0.08 0.9346
## Year1998 -0.04305 0.10149 -0.42 0.6715
## Year1999 0.11640 0.09294 1.25 0.2106
## Year2000 0.10434 0.09005 1.16 0.2467
## Year2001 0.10794 0.08851 1.22 0.2228
## Year2002 0.22104 0.10450 2.12 0.0346 *
## Year2003 0.17604 0.08646 2.04 0.0419 *
## Year2004 0.18665 0.09521 1.96 0.0501 .
## Year2005 0.12900 0.09472 1.36 0.1734
## Year2006 0.23593 0.08804 2.68 0.0074 **
```

```

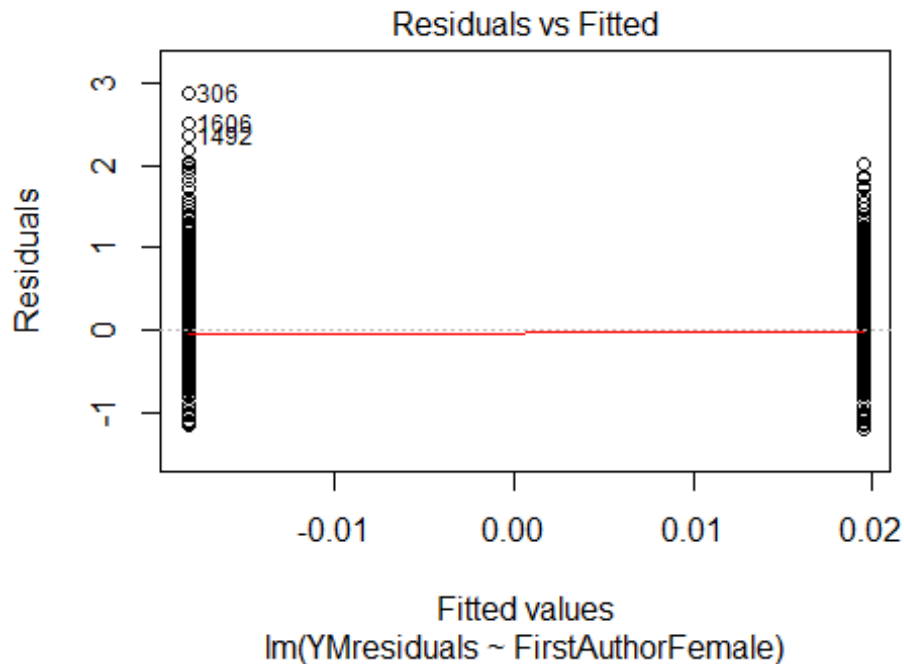
## Year2007      0.19939      0.08563      2.33      0.0200 *
## Year2008      0.07898      0.08542      0.92      0.3553
## Year2009      0.22199      0.08380      2.65      0.0082 **
## Year2010      0.16070      0.08428      1.91      0.0567 .
## Year2011      0.21362      0.09278      2.30      0.0214 *
## Year2012      0.16060      0.08423      1.91      0.0568 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.00424
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.165  0.873  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      6.58e-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: 1520"
## [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
##   29   35   22   36   39   32   43   41   49   65   82  107  164  171  190
## 2011 2012
##   205  264
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   27   18   32   30   26   40   35   41   62   68   91  147  148  169
## 2011 2012

```

```
## 181 226
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 25 27 18 31 30 25 40 33 39 61 66 91 146 141 164
## 2011 2012
## 178 225
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 75, df = 16, p-value = 1e-09
```

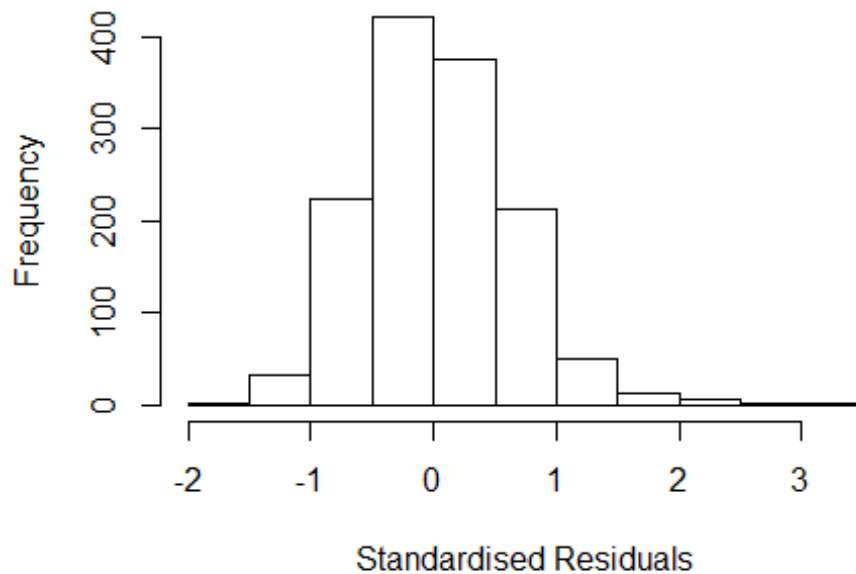


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.3, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 1.60452667044214"
## [1] "Male first author team size 2018 geometric mean: 1.53071802762622"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5900, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.58046648543001"
## [1] "Male last author team size 2018 geometric mean: 1.56724406173782"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5700, 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 2.315 1      1.521
## LastAuthorFemale  2.293 1      1.514
## UniqueAuthors     1.199 4      1.023
## Year              1.261 16     1.007
```

Residuals from first and last author and team size



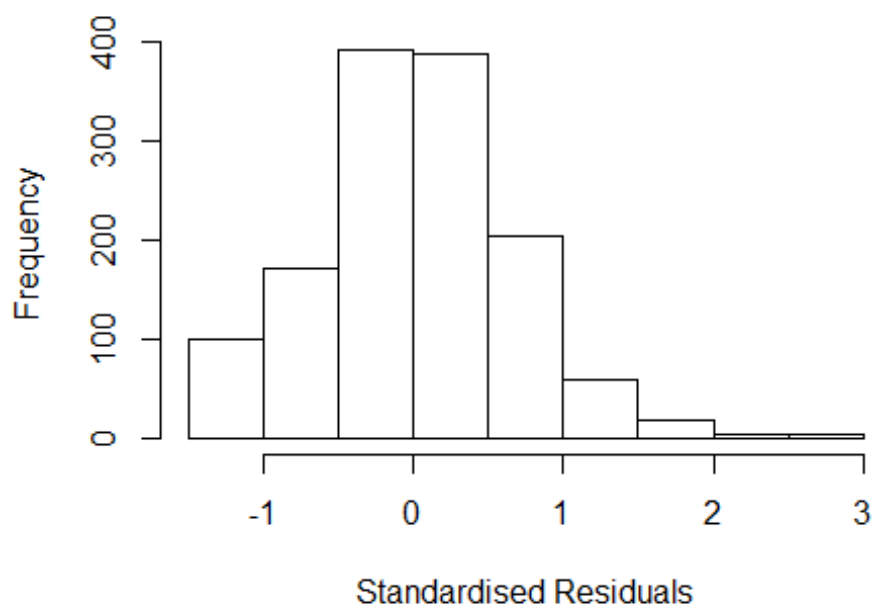
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 306   2442613368 4.011 2003    1213     2    3.044
## 1606 84859328881 3.575 2012    1203     3    2.692
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5116 -0.3760 -0.0118  0.4074  3.0438
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5908    0.0990     5.97 3.1e-09 ***
## FirstAuthorFemale1 0.1056    0.0507     2.08 0.03752 *
## LastAuthorFemale1 -0.0445    0.0505    -0.88 0.37812
## UniqueAuthors2    0.2532    0.0441     5.74 1.2e-08 ***
## UniqueAuthors3    0.5676    0.0684     8.30 2.5e-16 ***
## UniqueAuthors4    0.3367    0.1159     2.90 0.00374 **
## UniqueAuthors5    0.6350    0.1974     3.22 0.00133 **
## Year1997        -0.1957    0.1303    -1.50 0.13322
## Year1998        -0.0168    0.1418    -0.12 0.90548
```

```

## Year1999          0.1275      0.1458      0.87  0.38179
## Year2000          0.1944      0.1275      1.53  0.12740
## Year2001          0.3233      0.1545      2.09  0.03653 *
## Year2002          0.2750      0.1431      1.92  0.05477 .
## Year2003          0.3764      0.1300      2.89  0.00385 **
## Year2004          0.3133      0.1415      2.21  0.02697 *
## Year2005          0.2811      0.1167      2.41  0.01609 *
## Year2006          0.3186      0.1155      2.76  0.00590 **
## Year2007          0.4280      0.1166      3.67  0.00025 ***
## Year2008          0.2876      0.1082      2.66  0.00795 **
## Year2009          0.3319      0.1098      3.02  0.00255 **
## Year2010          0.2071      0.1086      1.91  0.05663 .
## Year2011          0.2874      0.1129      2.54  0.01107 *
## Year2012          0.2921      0.1142      2.56  0.01066 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0933
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 223 is an outlier with |weight| = 0 ( < 7.5e-05);
## 117 weights are ~= 1. The remaining 1222 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0125 0.8660 0.9530 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          7.46e-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 2.645 1 1.626
## LastAuthorFemale 2.639 1 1.624
## Year 1.115 16 1.003

```

Residuals from first and last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 306   2442613368 4.011 2003    1213     2    2.938
## 1492 84876523527 3.442 2012    3315     2    2.510
## 1606 84859328881 3.575 2012    1203     3    2.571
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.14793 -0.40842  0.00877  0.43095  2.93774
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6299    0.0901   6.99 4.2e-12 ***
## FirstAuthorFemale1 0.1177    0.0566   2.08 0.03779 *
## LastAuthorFemale1 -0.0716    0.0565  -1.27 0.20510
## Year1997        -0.2055    0.1268  -1.62 0.10536
## Year1998         0.0248    0.1395   0.18 0.85888
## Year1999         0.1348    0.1403   0.96 0.33676
## Year2000         0.2660    0.1294   2.06 0.04002 *
## Year2001         0.3704    0.1570   2.36 0.01848 *
## Year2002         0.3561    0.1430   2.49 0.01291 *
## Year2003         0.4433    0.1263   3.51 0.00046 ***
```

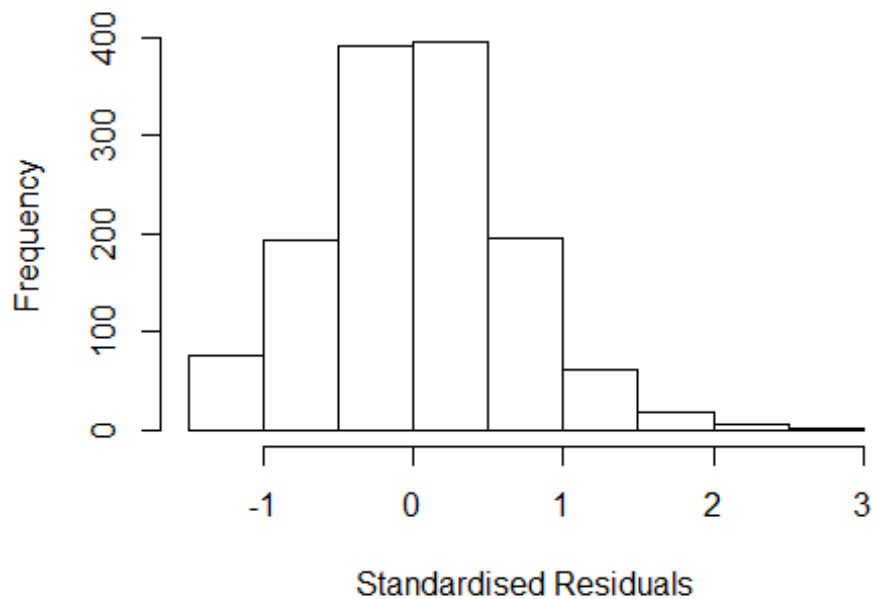


```

## Year2004          0.3938      0.1411      2.79  0.00533 **
## Year2005          0.3547      0.1094      3.24  0.00122 **
## Year2006          0.3989      0.1083      3.68  0.00024 ***
## Year2007          0.4720      0.1093      4.32  1.7e-05 ***
## Year2008          0.3546      0.1004      3.53  0.00043 ***
## Year2009          0.3937      0.1018      3.87  0.00011 ***
## Year2010          0.2771      0.1026      2.70  0.00701 **
## Year2011          0.3532      0.1067      3.31  0.00095 ***
## Year2012          0.3740      0.1054      3.55  0.00040 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0393, Adjusted R-squared:  0.0262
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 1235 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8700 0.9520 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      7.46e-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 1.065 1      1.032
## Year              1.065 16      1.002

```

Residuals from first author



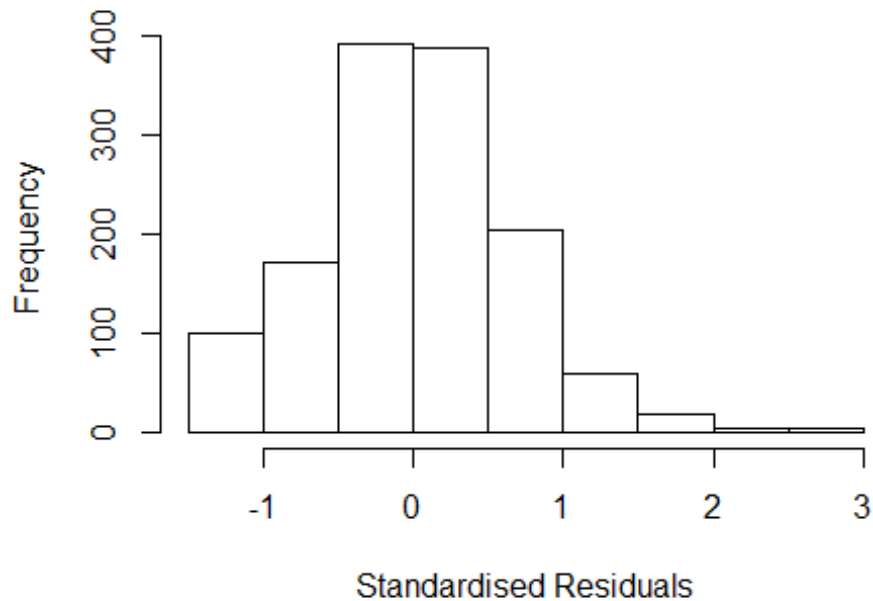
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 306   2442613368 4.011 2003    1213     2    2.938
## 1492 84876523527 3.442 2012    3315     2    2.510
## 1606 84859328881 3.575 2012    1203     3    2.571
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.15368 -0.40169  0.00442  0.42302  2.94355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6194    0.0898   6.90 8.3e-12 ***
## FirstAuthorFemale1 0.0622    0.0359   1.73 0.08378 .
## Year1997      -0.2009    0.1266  -1.59 0.11283
## Year1998       0.0273    0.1389   0.20 0.84428
## Year1999       0.1367    0.1400   0.98 0.32886
## Year2000       0.2702    0.1295   2.09 0.03716 *
## Year2001       0.3768    0.1578   2.39 0.01707 *
## Year2002       0.3598    0.1444   2.49 0.01283 *
## Year2003       0.4480    0.1266   3.54 0.00041 ***
## Year2004       0.3895    0.1414   2.75 0.00597 **
```

```

## Year2005          0.3563      0.1095      3.25  0.00117 **
## Year2006          0.3978      0.1086      3.66  0.00026 ***
## Year2007          0.4721      0.1093      4.32  1.7e-05 ***
## Year2008          0.3590      0.1004      3.58  0.00036 ***
## Year2009          0.3983      0.1016      3.92  9.3e-05 ***
## Year2010          0.2801      0.1025      2.73  0.00635 **
## Year2011          0.3583      0.1067      3.36  0.00080 ***
## Year2012          0.3763      0.1057      3.56  0.00039 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.629
## Multiple R-squared:  0.0379, Adjusted R-squared:  0.0255
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 223 is an outlier with |weight| <= 1.1e-05 ( < 7.5e-05);
## 94 weights are ~= 1. The remaining 1245 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0551 0.8690 0.9540 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      7.46e-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 1.062 1          1.031
## Year            1.062 16          1.002

```

Residuals from last author



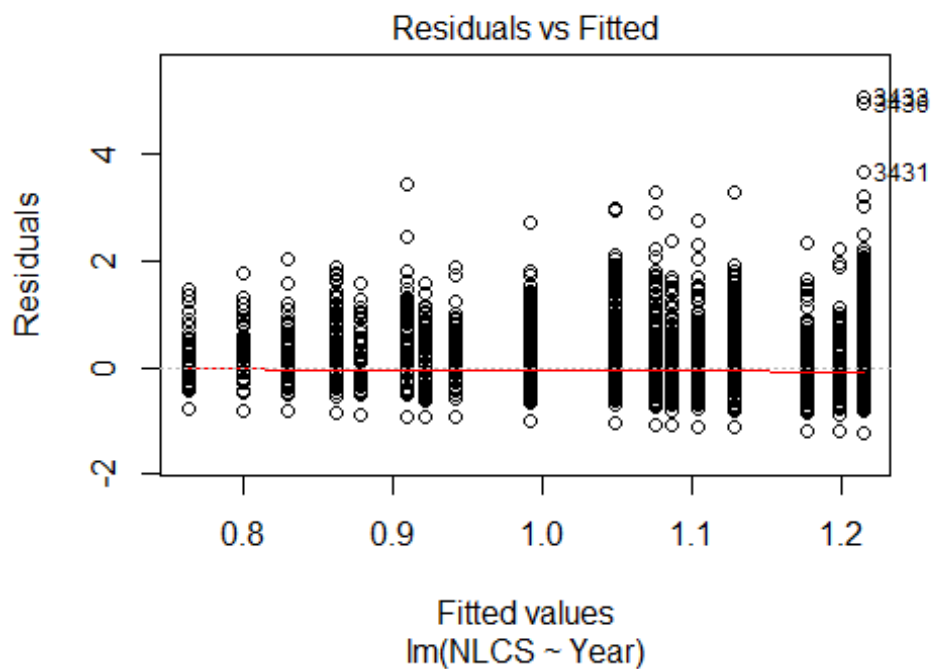
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 306   2442613368 4.011 2003    1213     2    2.938
## 1492 84876523527 3.442 2012    3315     2    2.510
## 1606 84859328881 3.575 2012    1203     3    2.571
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.13180 -0.40703  0.00272  0.42346  2.91904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6383    0.0913   6.99 4.2e-12 ***
## LastAuthorFemale1 0.0190    0.0359   0.53 0.59665
## Year1997       -0.2088    0.1261  -1.66 0.09803 .
## Year1998        0.0168    0.1390   0.12 0.90394
## Year1999        0.1235    0.1408   0.88 0.38068
## Year2000        0.2678    0.1305   2.05 0.04031 *
## Year2001        0.3757    0.1589   2.37 0.01817 *
## Year2002        0.3636    0.1453   2.50 0.01244 *
## Year2003        0.4536    0.1276   3.55 0.00039 ***
## Year2004        0.3956    0.1429   2.77 0.00572 **
```

```

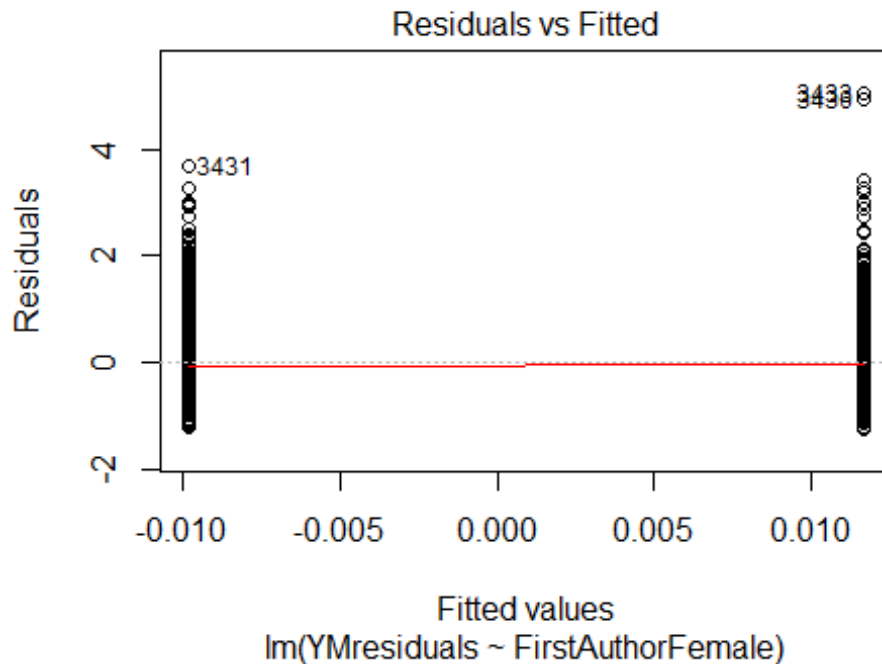
## Year2005          0.3566      0.1104      3.23  0.00126 **
## Year2006          0.3985      0.1096      3.64  0.00029 ***
## Year2007          0.4744      0.1106      4.29  1.9e-05 ***
## Year2008          0.3600      0.1014      3.55  0.00040 ***
## Year2009          0.4028      0.1026      3.93  9.0e-05 ***
## Year2010          0.2836      0.1032      2.75  0.00607 **
## Year2011          0.3634      0.1074      3.38  0.00074 ***
## Year2012          0.3822      0.1067      3.58  0.00035 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.631
## Multiple R-squared:  0.0358, Adjusted R-squared:  0.0234
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1252 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8720 0.9550 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      7.46e-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: 1340"
## [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
##   64   69   95   90  108   99  124  129  125  133  193  206  320  344  415
## 2011 2012
##   495  524
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 58 60 79 84 91 84 101 106 106 117 152 167 272 291 358
## 2011 2012
## 425 444
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 58 60 78 83 88 79 97 102 104 116 151 163 266 283 344
## 2011 2012
## 412 435
## [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 = 8e-14
```

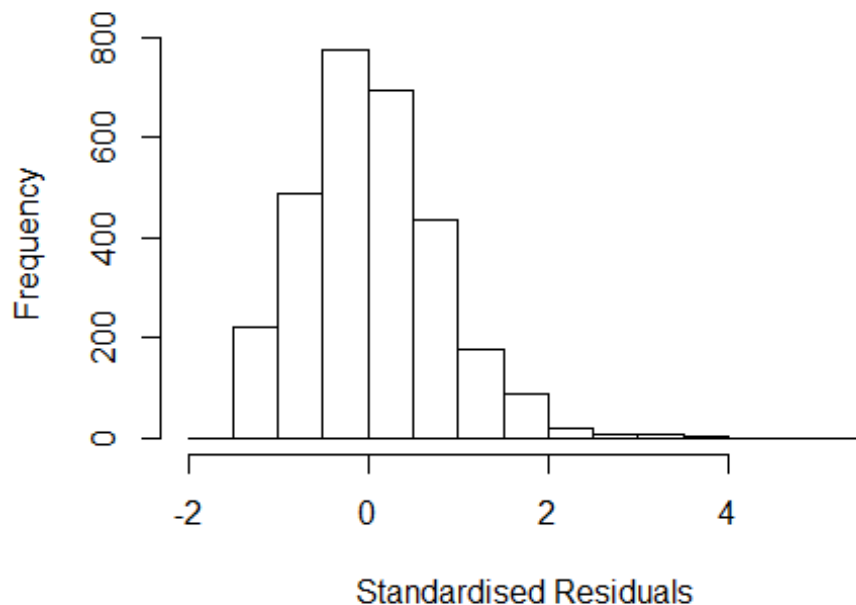


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.84, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 1.33356522075431"
## [1] "Male first author team size 2018 geometric mean: 1.22322312726859"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 35000, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.31534243960631"
## [1] "Male last author team size 2018 geometric mean: 1.24409786128961"
##
## 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 3.497 1      1.870
## LastAuthorFemale 3.426 1      1.851
## UniqueAuthors    1.202 4      1.023
## Year             1.147 16      1.004
```

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
## 710  34249125569 3.357 2003    1213      2    2.548
## 730  34247575975 4.335 2003    1213      2    3.526
## 1452 84998186713 3.843 2007    1213      2    2.806
## 1516 60950625388 3.992 2008    1213      2    3.003
## 1679 67649376492 4.345 2008    1213      2    3.356
## 1713 84997907764 3.992 2008    1213      2    2.875
## 1970 67650816432 4.398 2009    1213      2    3.170
## 2475 77956616722 3.699 2010    3316      1    2.794
## 2661 84860870363 4.032 2011    1213      2    3.086
## 2736 84555223668 3.991 2011    1202      2    3.045
## 3292 84866119495 4.214 2012    1213      2    2.979
## 3326 84865681764 3.712 2012    1213      2    2.687
## 3430 84861581198 3.712 2012    1213      2    2.646
## 3431 84861592197 4.875 2012    1213      2    3.850
## 3433 84861621527 6.265 2012    1213      2    5.030
## 3435 84861632269 4.410 2012    1213      2    3.344
## 3436 84861651000 6.164 2012    1213      2    4.929
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
```



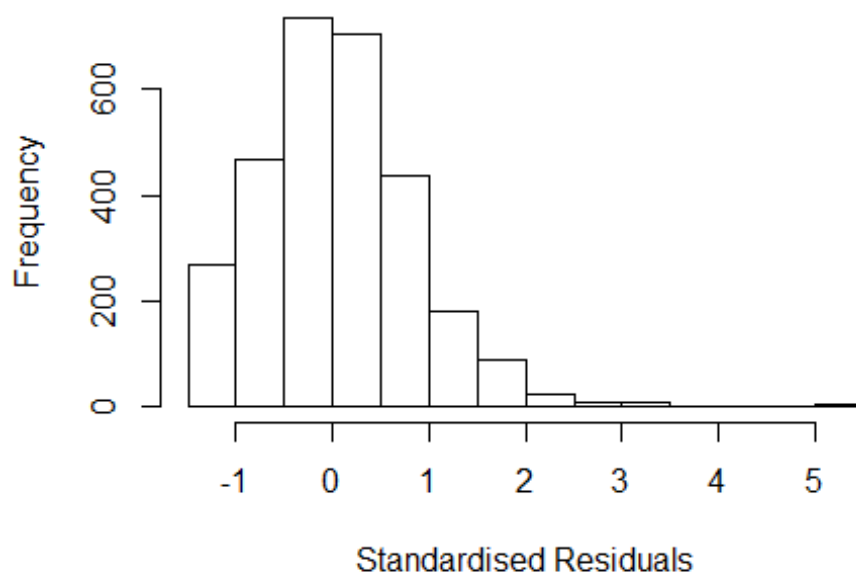
```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5351 -0.4870 -0.0122  0.5042  5.0302
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.81462    0.08621   9.45 < 2e-16 ***
## FirstAuthorFemale1 -0.03658    0.05316  -0.69  0.49145
## LastAuthorFemale1  0.07740    0.05271   1.47  0.14210
## UniqueAuthors2     0.16923    0.04391   3.85  0.00012 ***
## UniqueAuthors3     0.28411    0.08081   3.52  0.00045 ***
## UniqueAuthors4     0.24960    0.10401   2.40  0.01647 *
## UniqueAuthors5     0.54782    0.20449   2.68  0.00743 **
## Year1997          -0.11105    0.12218  -0.91  0.36349
## Year1998          -0.07919    0.11431  -0.69  0.48853
## Year1999           0.03789    0.11406   0.33  0.73975
## Year2000          -0.05116    0.11075  -0.46  0.64415
## Year2001           0.09270    0.10835   0.86  0.39233
## Year2002          -0.00537    0.12030  -0.04  0.96439
## Year2003          -0.04665    0.12039  -0.39  0.69841
## Year2004           0.30406    0.11669   2.61  0.00921 **
## Year2005           0.27508    0.11458   2.40  0.01642 *
## Year2006           0.20091    0.10464   1.92  0.05494 .
## Year2007           0.18179    0.10773   1.69  0.09163 .
## Year2008           0.13324    0.09743   1.37  0.17156
## Year2009           0.24456    0.09622   2.54  0.01109 *
## Year2010           0.09083    0.09428   0.96  0.33541
## Year2011           0.13188    0.09568   1.38  0.16820
## Year2012           0.21012    0.09707   2.16  0.03049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.758
## Multiple R-squared:  0.0333, Adjusted R-squared:  0.0259
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(2685,2687,2690)
## are outliers with |weight| = 0 ( < 3.4e-05);
## 233 weights are ~= 1. The remaining 2683 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002  0.8630  0.9530  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      3.43e-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 3.470 1      1.863
## LastAuthorFemale  3.369 1      1.835
## Year              1.079 16      1.002
```

Residuals from first and last author



```
## [1] "List of 17 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 710  34249125569 3.357 2003    1213     2    2.504
## 730  34247575975 4.335 2003    1213     2    3.482
## 1452 84998186713 3.843 2007    1213     2    2.773
## 1516 60950625388 3.992 2008    1213     2    2.958
## 1679 67649376492 4.345 2008    1213     2    3.311
## 1713 84997907764 3.992 2008    1213     2    3.003
## 1970 67650816432 4.398 2009    1213     2    3.302
## 2475 77956616722 3.699 2010    3316     1    2.753
## 2661 84860870363 4.032 2011    1213     2    3.054
## 2736 84555223668 3.991 2011    1202     2    3.013
```

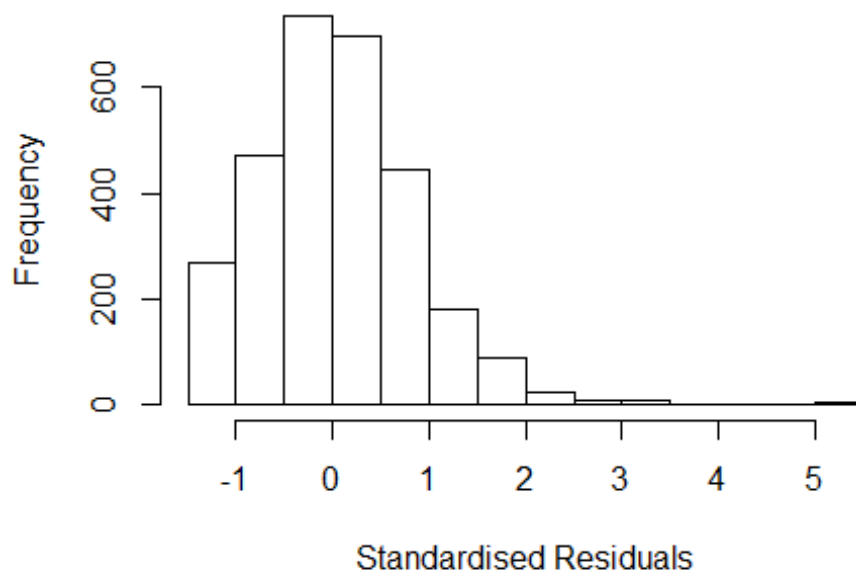
```

## 3292 84866119495 4.214 2012      1213      2      3.094
## 3326 84865681764 3.712 2012      1213      2      2.637
## 3430 84861581198 3.712 2012      1213      2      2.592
## 3431 84861592197 4.875 2012      1213      2      3.800
## 3433 84861621527 6.265 2012      1213      2      5.145
## 3435 84861632269 4.410 2012      1213      2      3.290
## 3436 84861651000 6.164 2012      1213      2      5.044
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1944 -0.5055 -0.0119  0.5123  5.1452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.84831    0.08556   9.91  <2e-16 ***
## FirstAuthorFemale1 -0.00388    0.05351  -0.07   0.9422
## LastAuthorFemale1  0.04915    0.05296   0.93   0.3534
## Year1997        -0.11804    0.12296  -0.96   0.3372
## Year1998        -0.08059    0.11536  -0.70   0.4849
## Year1999         0.03051    0.11404   0.27   0.7890
## Year2000        -0.06652    0.11035  -0.60   0.5467
## Year2001         0.09539    0.10972   0.87   0.3847
## Year2002        -0.01257    0.12039  -0.10   0.9168
## Year2003        -0.04104    0.12137  -0.34   0.7353
## Year2004         0.30086    0.11582   2.60   0.0094 **
## Year2005         0.27506    0.11286   2.44   0.0149 *
## Year2006         0.20181    0.10515   1.92   0.0550 .
## Year2007         0.17673    0.10767   1.64   0.1008
## Year2008         0.14077    0.09734   1.45   0.1482
## Year2009         0.24778    0.09618   2.58   0.0100 *
## Year2010         0.09756    0.09429   1.03   0.3009
## Year2011         0.12985    0.09560   1.36   0.1745
## Year2012         0.22621    0.09687   2.34   0.0196 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.763
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.0145
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(2685,2687,2690)
## are outliers with |weight| = 0 ( < 3.4e-05);
## 244 weights are ~ = 1. The remaining 2672 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.

```

```
## 0.0026 0.8560 0.9520 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          3.43e-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 1.062 1          1.030
## Year              1.062 16          1.002
```

Residuals from first author



```
## [1] "List of 17 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 710 34249125569 3.357 2003 1213 2 2.504
## 730 34247575975 4.335 2003 1213 2 3.482
## 1452 84998186713 3.843 2007 1213 2 2.773
```

```

## 1516 60950625388 3.992 2008      1213      2      2.958
## 1679 67649376492 4.345 2008      1213      2      3.311
## 1713 84997907764 3.992 2008      1213      2      3.003
## 1970 67650816432 4.398 2009      1213      2      3.302
## 2475 77956616722 3.699 2010     3316      1      2.753
## 2661 84860870363 4.032 2011      1213      2      3.054
## 2736 84555223668 3.991 2011     1202      2      3.013
## 3292 84866119495 4.214 2012      1213      2      3.094
## 3326 84865681764 3.712 2012      1213      2      2.637
## 3430 84861581198 3.712 2012      1213      2      2.592
## 3431 84861592197 4.875 2012      1213      2      3.800
## 3433 84861621527 6.265 2012      1213      2      5.145
## 3435 84861632269 4.410 2012      1213      2      3.290
## 3436 84861651000 6.164 2012      1213      2      5.044
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1899 -0.5078 -0.0148  0.5151  5.1495
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8531     0.0852   10.01  <2e-16 ***
## FirstAuthorFemale1  0.0374     0.0296    1.26  0.2062
## Year1997        -0.1196     0.1230   -0.97  0.3308
## Year1998        -0.0828     0.1153   -0.72  0.4729
## Year1999         0.0293     0.1140    0.26  0.7969
## Year2000        -0.0683     0.1103   -0.62  0.5359
## Year2001         0.0956     0.1098    0.87  0.3838
## Year2002        -0.0145     0.1203   -0.12  0.9038
## Year2003        -0.0412     0.1215   -0.34  0.7343
## Year2004         0.2994     0.1157    2.59  0.0097 **
## Year2005         0.2747     0.1128    2.44  0.0149 *
## Year2006         0.2019     0.1049    1.92  0.0545 .
## Year2007         0.1734     0.1074    1.61  0.1066
## Year2008         0.1385     0.0971    1.43  0.1541
## Year2009         0.2448     0.0959    2.55  0.0108 *
## Year2010         0.0953     0.0940    1.01  0.3112
## Year2011         0.1275     0.0954    1.34  0.1812
## Year2012         0.2250     0.0967    2.33  0.0200 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.763
## Multiple R-squared:  0.0202, Adjusted R-squared:  0.0145
## Convergence in 11 IRWLS iterations

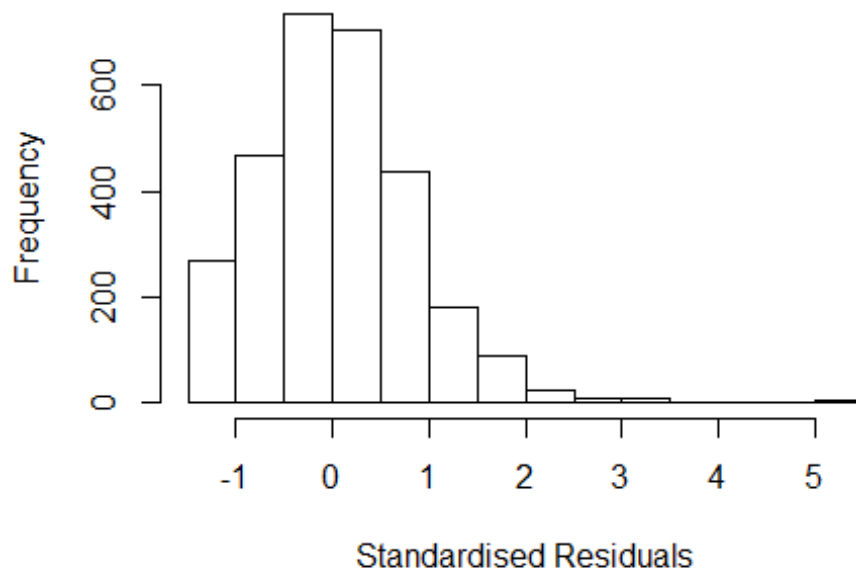
```

```

##
## Robustness weights:
## 3 observations c(2685,2687,2690)
## are outliers with |weight| = 0 ( < 3.4e-05);
## 238 weights are ~= 1. The remaining 2678 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0024 0.8550 0.9530 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           3.43e-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 1.031 1           1.015
## Year           1.031 16           1.001

```

Residuals from last author



```
## [1] "List of 17 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 710  34249125569 3.357 2003    1213     2    2.504
## 730  34247575975 4.335 2003    1213     2    3.482
## 1452 84998186713 3.843 2007    1213     2    2.773
## 1516 60950625388 3.992 2008    1213     2    2.958
## 1679 67649376492 4.345 2008    1213     2    3.311
## 1713 84997907764 3.992 2008    1213     2    3.003
## 1970 67650816432 4.398 2009    1213     2    3.302
## 2475 77956616722 3.699 2010    3316     1    2.753
## 2661 84860870363 4.032 2011    1213     2    3.054
## 2736 84555223668 3.991 2011    1202     2    3.013
## 3292 84866119495 4.214 2012    1213     2    3.094
## 3326 84865681764 3.712 2012    1213     2    2.637
## 3430 84861581198 3.712 2012    1213     2    2.592
## 3431 84861592197 4.875 2012    1213     2    3.800
## 3433 84861621527 6.265 2012    1213     2    5.145
## 3435 84861632269 4.410 2012    1213     2    3.290
## 3436 84861651000 6.164 2012    1213     2    5.044
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -1.1948 -0.5056 -0.0116 0.5122 5.1445
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8485     0.0855   9.92  <2e-16 ***
## LastAuthorFemale1 0.0458     0.0293   1.56  0.1179
## Year1997         -0.1182     0.1228  -0.96  0.3360
## Year1998         -0.0808     0.1153  -0.70  0.4835
## Year1999          0.0303     0.1138   0.27  0.7904
## Year2000         -0.0668     0.1101  -0.61  0.5442
## Year2001          0.0952     0.1097   0.87  0.3853
## Year2002         -0.0126     0.1203  -0.10  0.9168
## Year2003         -0.0409     0.1213  -0.34  0.7363
## Year2004          0.3005     0.1157   2.60  0.0094 **
## Year2005          0.2748     0.1127   2.44  0.0148 *
## Year2006          0.2016     0.1049   1.92  0.0548 .
## Year2007          0.1764     0.1072   1.65  0.0999 .
## Year2008          0.1405     0.0969   1.45  0.1470
## Year2009          0.2474     0.0957   2.59  0.0098 **
## Year2010          0.0971     0.0939   1.03  0.3008
## Year2011          0.1296     0.0949   1.37  0.1723
## Year2012          0.2262     0.0966   2.34  0.0192 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.765
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.0148
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 3 observations c(2685,2687,2690)
## are outliers with |weight| = 0 ( < 3.4e-05);
## 244 weights are ~ = 1. The remaining 2672 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0033 0.8570 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           3.43e-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"

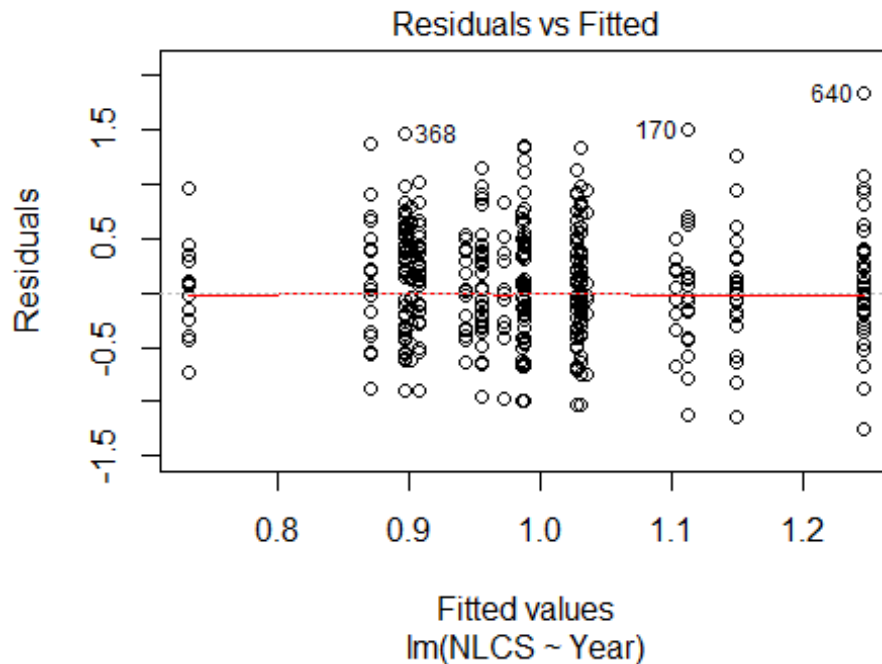
```



```

## seed : int(0)
## [1] "Sample size for the above analysis: 2919"
## [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
##    33    22    22    15    24    19    32    28    20    16    61    68    54    57    53
## 2011 2012
##    43    51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    23    17    17    11    16     8    20    21    15    11    51    63    45    43    46
## 2011 2012
##    35    37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    22    16    17    11    16     8    17    20    15    10    48    62    43    41    43
## 2011 2012
##    34    36
## [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 = 0.4

```



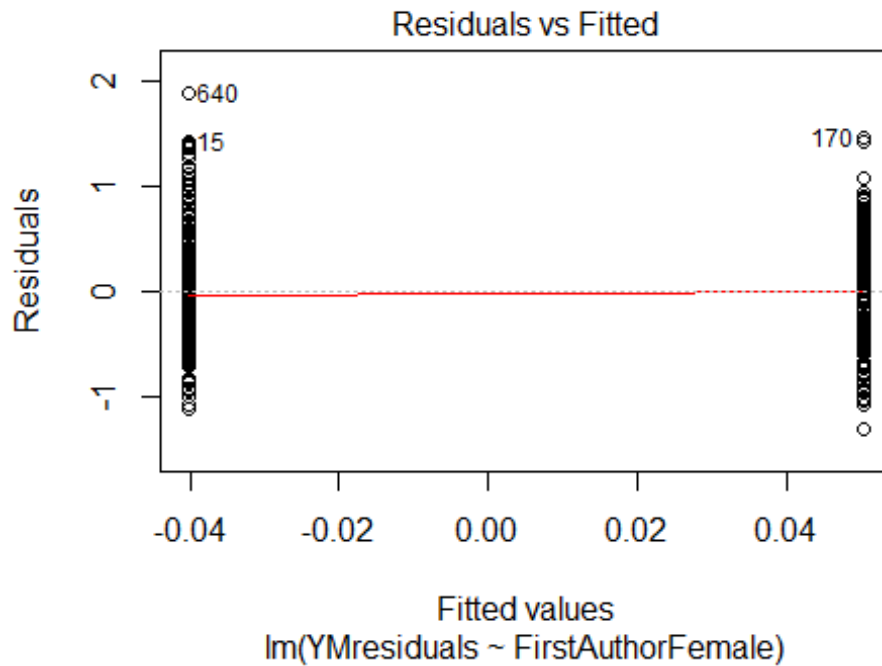
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.68, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 1.51398420474934"
## [1] "Male first author team size 2018 geometric mean: 1.67806061608904"

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

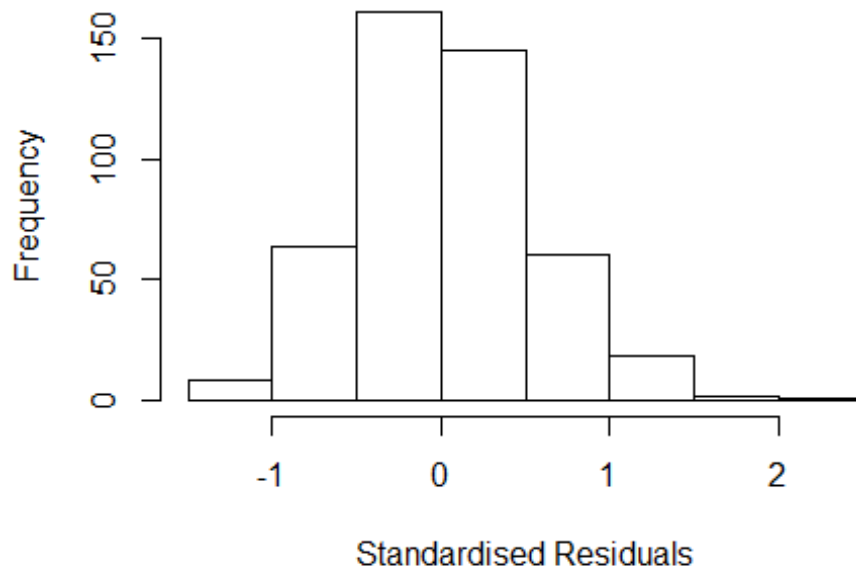
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 370, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.6424692368679"
## [1] "Male last author team size 2018 geometric mean: 1.48976673833581"

## 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 = 0.5
## 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 2.157 1          1.469
## LastAuthorFemale 2.143 1          1.464
## UniqueAuthors    2.160 4          1.101
## Year              2.788 16         1.033
```

Residuals from first and last author and team size



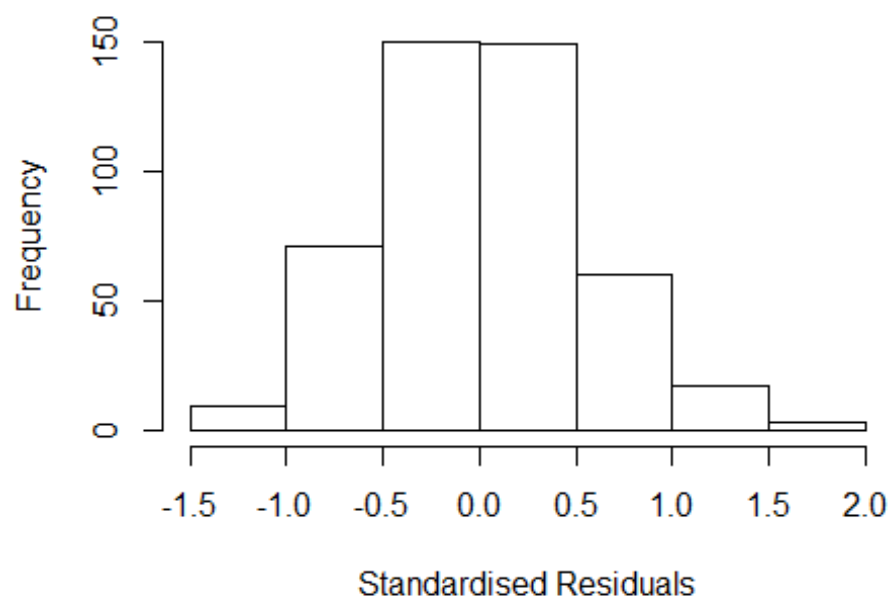
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34169 -0.35600 -0.00492 0.34865 2.03211
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7172 0.1302 5.51 6.1e-08 ***
## FirstAuthorFemale1 0.1058 0.0726 1.46 0.1461
## LastAuthorFemale1 0.0647 0.0732 0.88 0.3775
## UniqueAuthors2 0.1214 0.0614 1.98 0.0488 *
## UniqueAuthors3 0.1548 0.0887 1.75 0.0816 .
## UniqueAuthors4 0.5693 0.2129 2.67 0.0078 **
## UniqueAuthors5 0.1352 0.1233 1.10 0.2735
## Year1997 0.1285 0.1822 0.71 0.4809
## Year1998 -0.0579 0.1864 -0.31 0.7563
## Year1999 0.0488 0.2178 0.22 0.8228
```

```

## Year2000      0.0768      0.1748      0.44      0.6607
## Year2001      0.1886      0.2698      0.70      0.4849
## Year2002      0.2309      0.1931      1.20      0.2324
## Year2003      0.2790      0.1640      1.70      0.0896 .
## Year2004      0.1373      0.1601      0.86      0.3915
## Year2005      0.2308      0.1616      1.43      0.1540
## Year2006      0.1850      0.1445      1.28      0.2011
## Year2007      0.0565      0.1435      0.39      0.6941
## Year2008      0.0758      0.1445      0.52      0.5999
## Year2009      0.1879      0.1614      1.16      0.2451
## Year2010      0.0439      0.1602      0.27      0.7843
## Year2011      0.3326      0.1754      1.90      0.0586 .
## Year2012      0.0101      0.1587      0.06      0.9493
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.527
## Multiple R-squared:  0.0766, Adjusted R-squared:  0.03
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 422 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  0.882  0.955   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      2.18e-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 2.083 1      1.443
## LastAuthorFemale 2.004 1      1.416
## Year      1.347 16      1.009

```

Residuals from first and last author



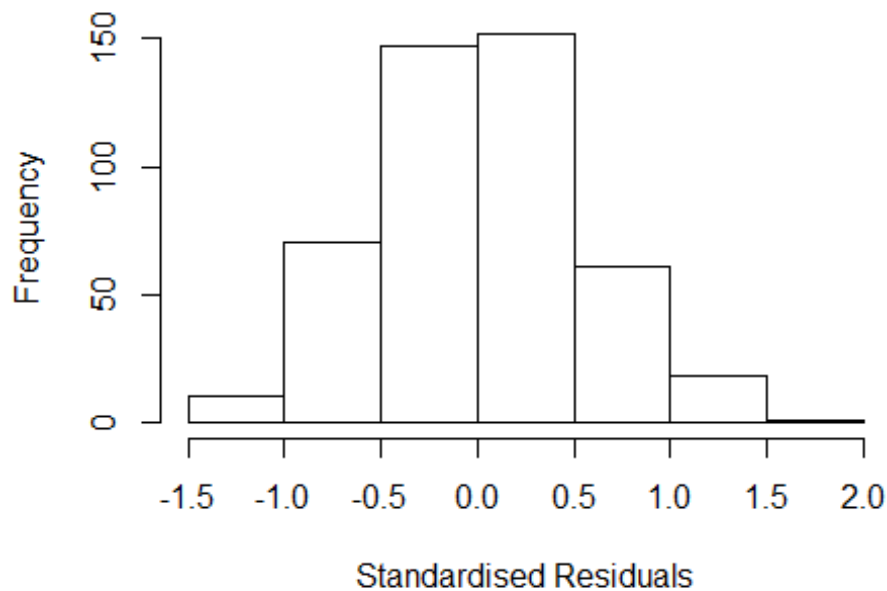
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27557 -0.35835 -0.00688 0.35677 1.96614
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7640 0.1358 5.63 3.3e-08 ***
## FirstAuthorFemale1 0.0982 0.0735 1.34 0.182
## LastAuthorFemale1 0.0615 0.0730 0.84 0.400
## Year1997 0.1326 0.1897 0.70 0.485
## Year1998 -0.0894 0.1883 -0.47 0.635
## Year1999 0.1364 0.2044 0.67 0.505
## Year2000 0.0755 0.1776 0.43 0.671
## Year2001 0.1602 0.2811 0.57 0.569
## Year2002 0.2285 0.2005 1.14 0.255
## Year2003 0.2740 0.1720 1.59 0.112
## Year2004 0.1172 0.1658 0.71 0.480
## Year2005 0.2515 0.1694 1.48 0.138
```

```

## Year2006          0.2061      0.1513      1.36      0.174
## Year2007          0.0703      0.1500      0.47      0.639
## Year2008          0.0926      0.1517      0.61      0.542
## Year2009          0.2121      0.1672      1.27      0.205
## Year2010          0.0562      0.1670      0.34      0.736
## Year2011          0.3518      0.1754      2.01      0.045 *
## Year2012          0.0575      0.1664      0.35      0.730
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0534, Adjusted R-squared:  0.0147
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 416 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.149  0.872  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      2.18e-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 1.192 1      1.092
## Year              1.192 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
## -1.2613 -0.3561 0.0013 0.3681 1.9613
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7707 0.1378 5.59 4e-08 ***
## FirstAuthorFemale1 0.1406 0.0557 2.53 0.012 *
## Year1997 0.1361 0.1913 0.71 0.477
## Year1998 -0.0918 0.1886 -0.49 0.627
## Year1999 0.1394 0.2036 0.68 0.494
## Year2000 0.0752 0.1797 0.42 0.676
## Year2001 0.1686 0.2843 0.59 0.553
## Year2002 0.2361 0.2015 1.17 0.242
## Year2003 0.2713 0.1725 1.57 0.117
## Year2004 0.1181 0.1674 0.71 0.481
## Year2005 0.2650 0.1697 1.56 0.119
## Year2006 0.2061 0.1529 1.35 0.178
```

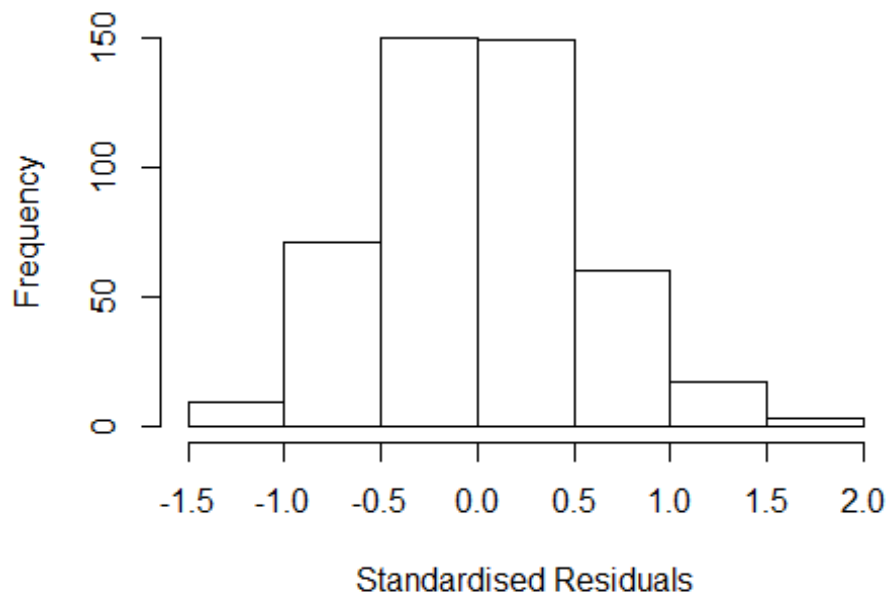


```

## Year2007          0.0688      0.1519      0.45      0.651
## Year2008          0.0977      0.1532      0.64      0.524
## Year2009          0.2165      0.1687      1.28      0.200
## Year2010          0.0610      0.1698      0.36      0.720
## Year2011          0.3500      0.1777      1.97      0.050 *
## Year2012          0.0658      0.1663      0.40      0.692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0516, Adjusted R-squared:  0.015
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 416 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.153  0.868  0.948  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      2.18e-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 1.138 1      1.067
## Year              1.138 16      1.004

```

Residuals from last author



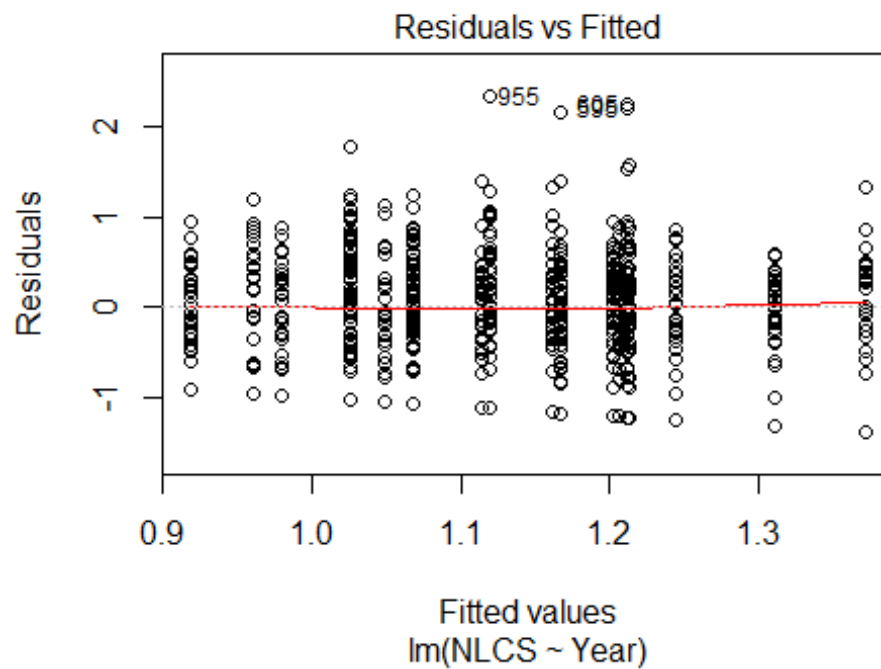
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27182 -0.35572 -0.00808 0.36955 1.93773
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7802 0.1321 5.91 7e-09 ***
## LastAuthorFemale1 0.1276 0.0552 2.31 0.021 *
## Year1997 0.1216 0.1851 0.66 0.512
## Year1998 -0.0977 0.1840 -0.53 0.596
## Year1999 0.1350 0.2034 0.66 0.507
## Year2000 0.0702 0.1761 0.40 0.690
## Year2001 0.1556 0.2682 0.58 0.562
## Year2002 0.2107 0.1969 1.07 0.285
## Year2003 0.2727 0.1717 1.59 0.113
## Year2004 0.1137 0.1633 0.70 0.487
## Year2005 0.2320 0.1654 1.40 0.162
## Year2006 0.2094 0.1491 1.40 0.161
```

```

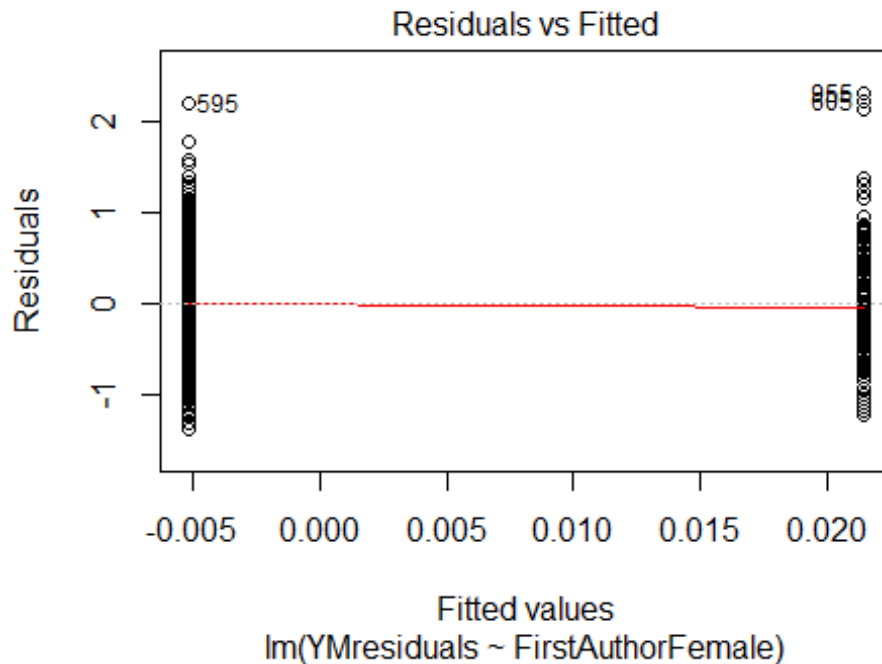
## Year2007          0.0719      0.1484      0.48      0.628
## Year2008          0.0921      0.1490      0.62      0.537
## Year2009          0.2040      0.1649      1.24      0.217
## Year2010          0.0612      0.1636      0.37      0.708
## Year2011          0.3641      0.1723      2.11      0.035 *
## Year2012          0.0567      0.1656      0.34      0.732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.049, Adjusted R-squared:  0.0124
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 414 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.177  0.880  0.950  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      2.18e-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: 459"
## [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
##   35   33   39   41   42   38   45   38   33   33   44   61   70   77   115
## 2011 2012
##   81   90
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   31   31   34   36   31   36   28   26   27   37   56   63   68   101
## 2011 2012

```

```
## 68 72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 31 31 30 33 36 29 35 26 26 27 36 53 60 67 99
## 2011 2012
## 66 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 = 26, df = 16, p-value = 0.06
```

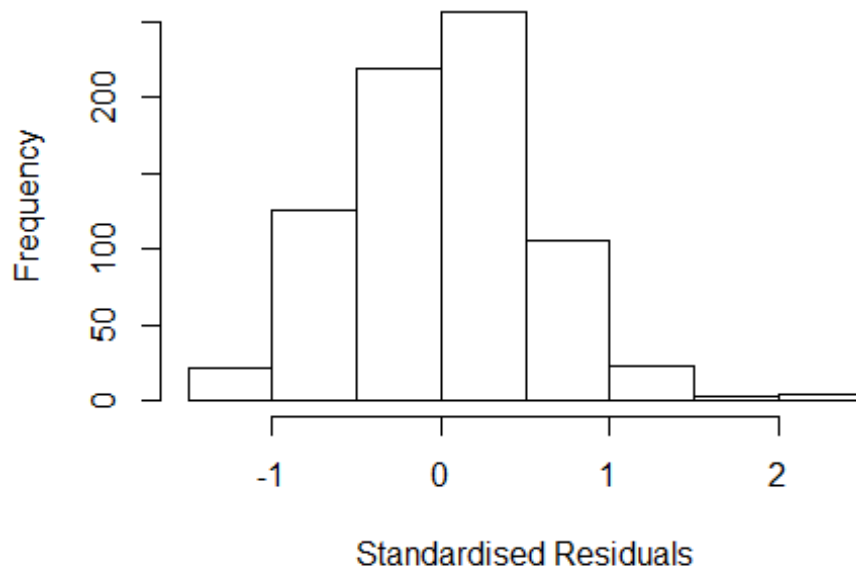


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 1.58788592588649"
## [1] "Male first author team size 2018 geometric mean: 1.3610605475846"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.46830747284373"
## [1] "Male last author team size 2018 geometric mean: 1.72467358728587"
##
## 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 1.991 1          1.411
## LastAuthorFemale  1.951 1          1.397
## UniqueAuthors    1.585 4          1.059
## Year             1.565 16          1.014
```

Residuals from first and last author and team size



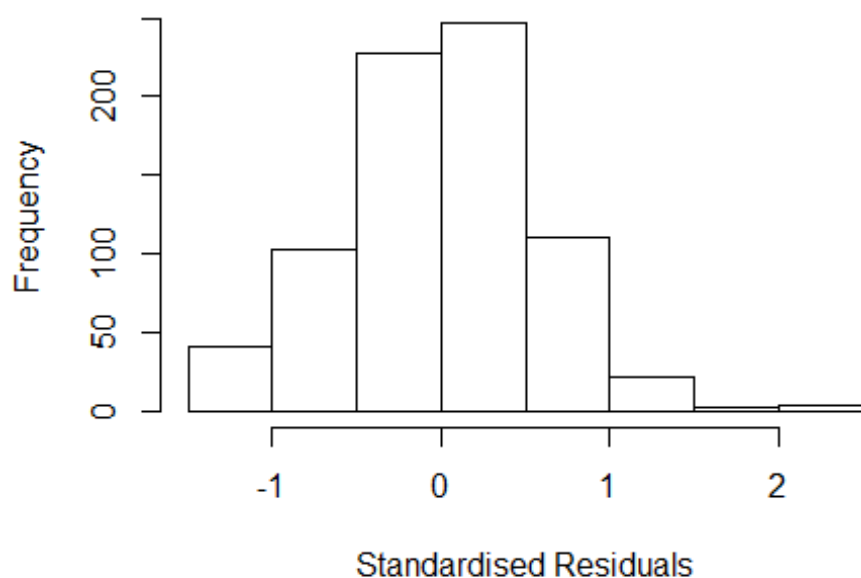
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4872 -0.3769 0.0213 0.3723 2.4323
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9033 0.1541 5.86 6.9e-09 ***
## FirstAuthorFemale1 0.0333 0.0737 0.45 0.65194
## LastAuthorFemale1 -0.0299 0.0676 -0.44 0.65911
## UniqueAuthors2 0.3333 0.0490 6.80 2.1e-11 ***
## UniqueAuthors3 0.3333 0.0809 4.12 4.2e-05 ***
## UniqueAuthors4 0.4507 0.1292 3.49 0.00051 ***
## UniqueAuthors5 0.4567 0.1299 3.52 0.00046 ***
## Year1997 0.1757 0.1725 1.02 0.30884
## Year1998 0.0244 0.1784 0.14 0.89125
## Year1999 -0.0390 0.1652 -0.24 0.81342
```

```

## Year2000          0.3364      0.1561      2.15  0.03152 *
## Year2001          0.1490      0.1767      0.84  0.39950
## Year2002          0.2471      0.1646      1.50  0.13365
## Year2003          0.1518      0.1737      0.87  0.38246
## Year2004         -0.0401      0.1874     -0.21  0.83081
## Year2005          0.3671      0.1910      1.92  0.05497 .
## Year2006          0.0688      0.1615      0.43  0.67019
## Year2007          0.2134      0.1588      1.34  0.17951
## Year2008          0.1264      0.1564      0.81  0.41906
## Year2009          0.1149      0.1520      0.76  0.45012
## Year2010         -0.0101      0.1539     -0.07  0.94780
## Year2011          0.0663      0.1548      0.43  0.66846
## Year2012          0.0531      0.1661      0.32  0.74945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.112, Adjusted R-squared:  0.0855
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 697 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0242 0.8760 0.9540 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          1.32e-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 1.959 1          1.400
## LastAuthorFemale  2.015 1          1.419
## Year              1.196 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
## -1.4336 -0.3965 0.0122 0.3989 2.3456
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98863 0.15848 6.24 7.5e-10 ***
## FirstAuthorFemale1 0.11237 0.07839 1.43 0.152
## LastAuthorFemale1 -0.12984 0.07140 -1.82 0.069 .
## Year1997 0.18597 0.17962 1.04 0.301
## Year1998 -0.00794 0.18329 -0.04 0.965
## Year1999 -0.04647 0.16966 -0.27 0.784
## Year2000 0.35079 0.16007 2.19 0.029 *
## Year2001 0.21584 0.18616 1.16 0.247
## Year2002 0.27224 0.17167 1.59 0.113
## Year2003 0.24344 0.18037 1.35 0.178
## Year2004 0.03560 0.18569 0.19 0.848
## Year2005 0.46241 0.18973 2.44 0.015 *
```

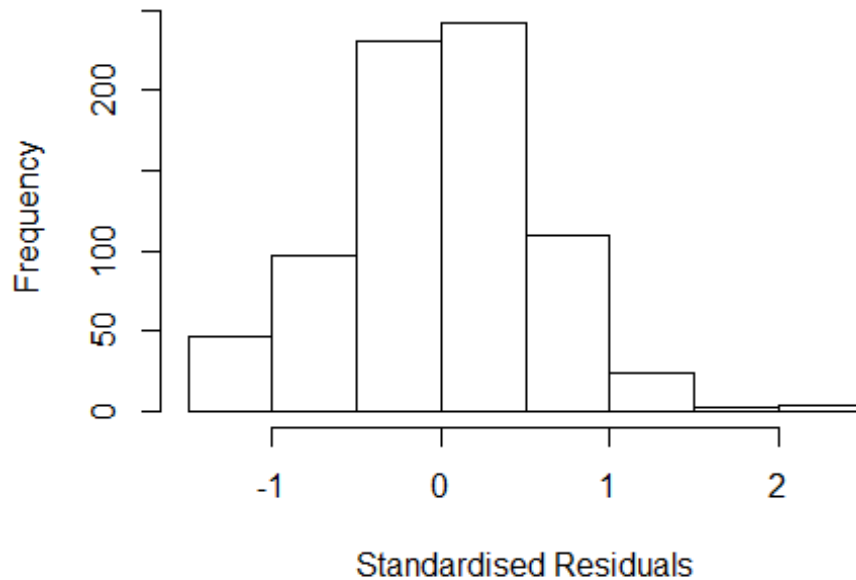


```

## Year2006          0.13052      0.16727      0.78      0.435
## Year2007          0.26530      0.16491      1.61      0.108
## Year2008          0.15966      0.16210      0.98      0.325
## Year2009          0.16708      0.16005      1.04      0.297
## Year2010          0.02592      0.16198      0.16      0.873
## Year2011          0.09661      0.16240      0.59      0.552
## Year2012          0.12178      0.17160      0.71      0.478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0425, Adjusted R-squared:  0.0192
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 695 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.074  0.879   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      1.32e-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 1.076 1          1.037
## Year              1.076 16          1.002

```

Residuals from first author



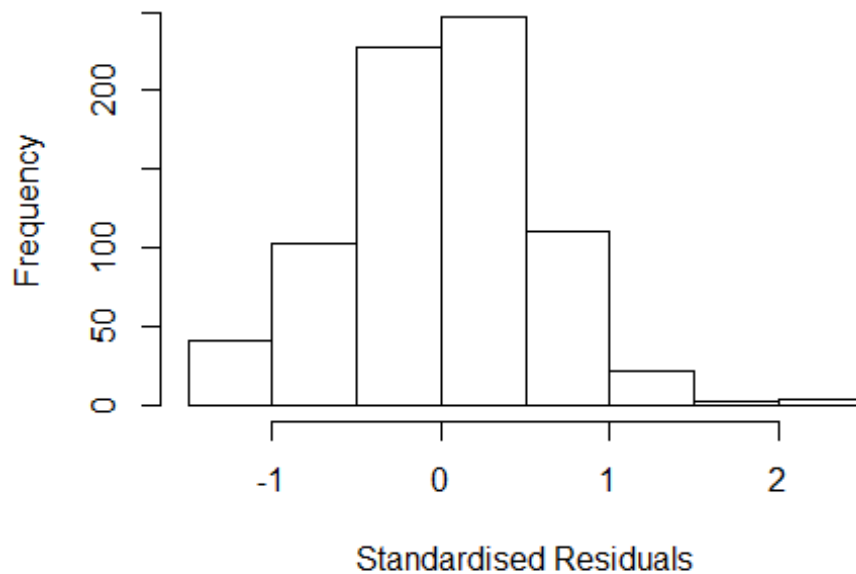
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4278 -0.3904 0.0154 0.4038 2.3668
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96217 0.15751 6.11 1.6e-09 ***
## FirstAuthorFemale1 0.01073 0.05749 0.19 0.852
## Year1997 0.19335 0.17937 1.08 0.281
## Year1998 -0.00557 0.18360 -0.03 0.976
## Year1999 -0.03978 0.17006 -0.23 0.815
## Year2000 0.36054 0.15992 2.25 0.024 *
## Year2001 0.22287 0.18894 1.18 0.239
## Year2002 0.28252 0.17237 1.64 0.102
## Year2003 0.26335 0.17726 1.49 0.138
## Year2004 0.06258 0.18438 0.34 0.734
## Year2005 0.45488 0.19011 2.39 0.017 *
## Year2006 0.14831 0.16658 0.89 0.374
```

```

## Year2007          0.26732      0.16508      1.62      0.106
## Year2008          0.17186      0.16241      1.06      0.290
## Year2009          0.17840      0.15988      1.12      0.265
## Year2010          0.03524      0.16216      0.22      0.828
## Year2011          0.09913      0.16267      0.61      0.542
## Year2012          0.12704      0.17231      0.74      0.461
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0391, Adjusted R-squared:  0.017
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 698 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0657 0.8810 0.9520 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      1.32e-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 1.113 1      1.055
## Year      1.113 16      1.003

```

Residuals from last author



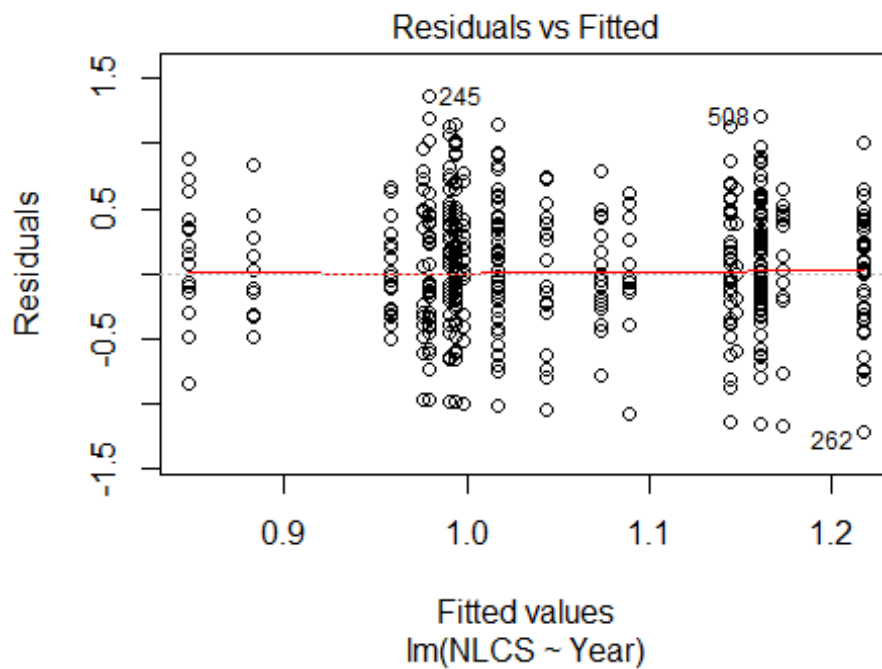
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.419 -0.385 0.010 0.406 2.315
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0216 0.1553 6.58 9e-11 ***
## LastAuthorFemale1 -0.0525 0.0540 -0.97 0.331
## Year1997 0.1859 0.1802 1.03 0.303
## Year1998 -0.0109 0.1831 -0.06 0.953
## Year1999 -0.0548 0.1690 -0.32 0.746
## Year2000 0.3538 0.1599 2.21 0.027 *
## Year2001 0.2193 0.1871 1.17 0.242
## Year2002 0.2659 0.1714 1.55 0.121
## Year2003 0.2545 0.1779 1.43 0.153
## Year2004 0.0418 0.1857 0.23 0.822
## Year2005 0.4499 0.1892 2.38 0.018 *
## Year2006 0.1310 0.1671 0.78 0.433
```

```

## Year2007          0.2560      0.1646      1.56      0.120
## Year2008          0.1620      0.1622      1.00      0.318
## Year2009          0.1674      0.1596      1.05      0.294
## Year2010          0.0259      0.1619      0.16      0.873
## Year2011          0.0885      0.1621      0.55      0.585
## Year2012          0.1196      0.1715      0.70      0.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0401, Adjusted R-squared:  0.0181
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 59 weights are ~= 1. The remaining 698 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0833 0.8800 0.9520 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      1.32e-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: 757"
## [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
##   18   11   22   10   16   21   23   20   19   16   40   41   54   47   52
## 2011 2012
##   74   84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   11   21    7   12   18   21   19   17   15   37   38   49   42   49
## 2011 2012

```

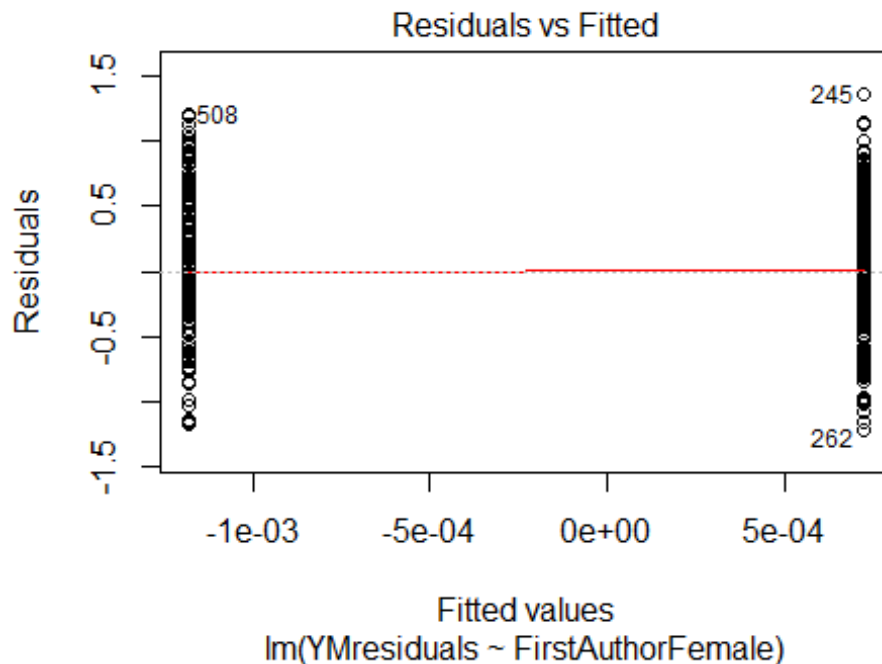
```
## 70 78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 11 18 7 12 16 21 19 16 15 36 36 45 39 47
## 2011 2012
## 68 77
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: 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 = 2.1, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 2.26328179059281"
## [1] "Male first author team size 2018 geometric mean: 2.37056619252531"
## 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 = 540, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.33254853810268"
## [1] "Male last author team size 2018 geometric mean: 2.2445783541436"

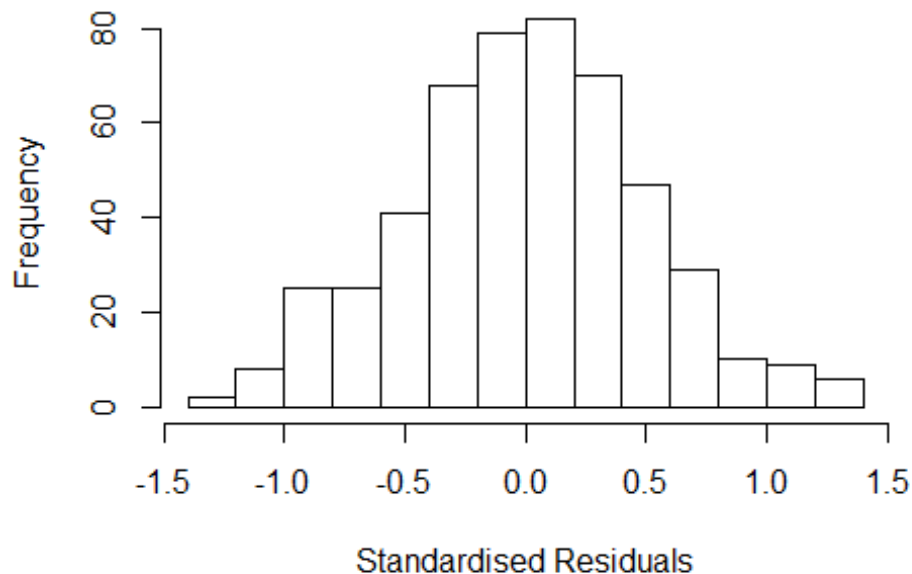
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 530, 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	1.374	1	1.172
LastAuthorFemale	1.323	1	1.150
UniqueAuthors	2.032	4	1.093
Year	2.432	16	1.028

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
## -1.33627 -0.30643 0.00216 0.31881 1.38881
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98919 0.14455 6.84 2.4e-11 ***
## FirstAuthorFemale1 -0.03174 0.05485 -0.58 0.5631
## LastAuthorFemale1 0.08451 0.05145 1.64 0.1011
## UniqueAuthors2 0.18497 0.05745 3.22 0.0014 **
## UniqueAuthors3 0.27925 0.06807 4.10 4.8e-05 ***
## UniqueAuthors4 0.26257 0.11308 2.32 0.0207 *
## UniqueAuthors5 0.16659 0.09379 1.78 0.0763 .
## Year1997 -0.26320 0.19117 -1.38 0.1692
## Year1998 -0.19925 0.15218 -1.31 0.1911
## Year1999 0.03207 0.22136 0.14 0.8849
```

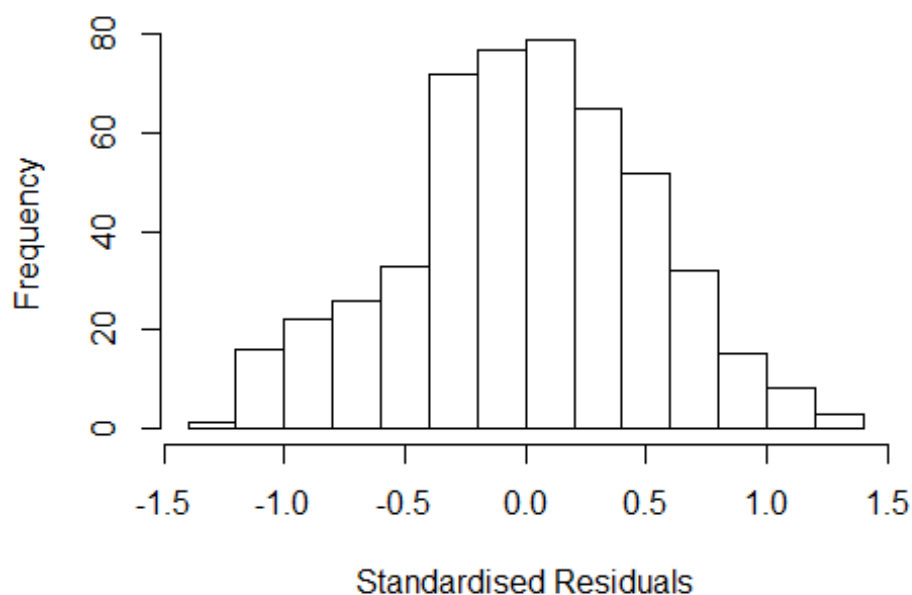


```

## Year2000      0.00324    0.18626    0.02    0.9861
## Year2001     -0.10228    0.16809   -0.61    0.5432
## Year2002     -0.20068    0.18385   -1.09    0.2756
## Year2003     -0.14526    0.19766   -0.73    0.4628
## Year2004     -0.07902    0.19012   -0.42    0.6779
## Year2005      0.03442    0.20067    0.17    0.8639
## Year2006     -0.16171    0.17389   -0.93    0.3529
## Year2007      0.04730    0.16333    0.29    0.7722
## Year2008     -0.17835    0.16805   -1.06    0.2891
## Year2009      0.00199    0.16568    0.01    0.9904
## Year2010     -0.16878    0.16502   -1.02    0.3069
## Year2011     -0.00400    0.16610   -0.02    0.9808
## Year2012     -0.18833    0.15886   -1.19    0.2364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.0909, Adjusted R-squared:  0.0491
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.386  0.863  0.956  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      2.00e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.312 1      1.145
## LastAuthorFemale  1.230 1      1.109
## Year              1.316 16      1.009

```

Residuals from first and last author



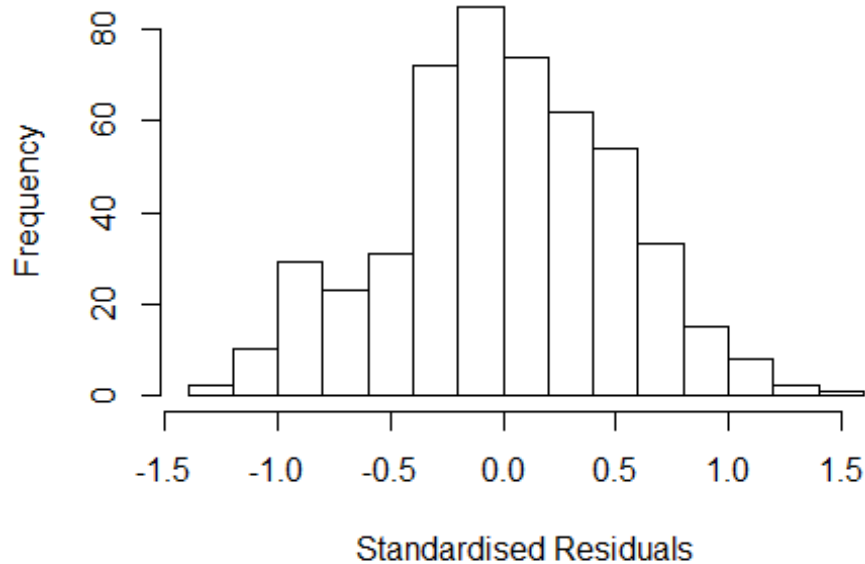
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24832 -0.31654  0.00559  0.34192  1.38868
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0446     0.1471    7.10 4.5e-12 ***
## FirstAuthorFemale1 -0.0274     0.0551   -0.50   0.62
## LastAuthorFemale1  0.0809     0.0515    1.57   0.12
## Year1997          -0.2003     0.1850   -1.08   0.28
## Year1998          -0.1653     0.1545   -1.07   0.29
## Year1999           0.0849     0.2126    0.40   0.69
## Year2000           0.0356     0.1879    0.19   0.85
## Year2001          -0.0451     0.1696   -0.27   0.79
## Year2002          -0.2032     0.1879   -1.08   0.28
## Year2003          -0.0998     0.2206   -0.45   0.65
## Year2004          -0.0533     0.1960   -0.27   0.79
## Year2005           0.1455     0.1981    0.73   0.46
```

```

## Year2006          -0.1439      0.1767    -0.81      0.42
## Year2007           0.1501      0.1654      0.91      0.36
## Year2008          -0.0959      0.1678    -0.57      0.57
## Year2009           0.0563      0.1687      0.33      0.74
## Year2010          -0.0778      0.1688    -0.46      0.65
## Year2011           0.0764      0.1609      0.48      0.63
## Year2012          -0.1028      0.1580    -0.65      0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0453, Adjusted R-squared:  0.00967
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 454 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.410  0.862  0.950  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      2.00e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.208 1          1.099
## Year              1.208 16          1.006

```

Residuals from first author



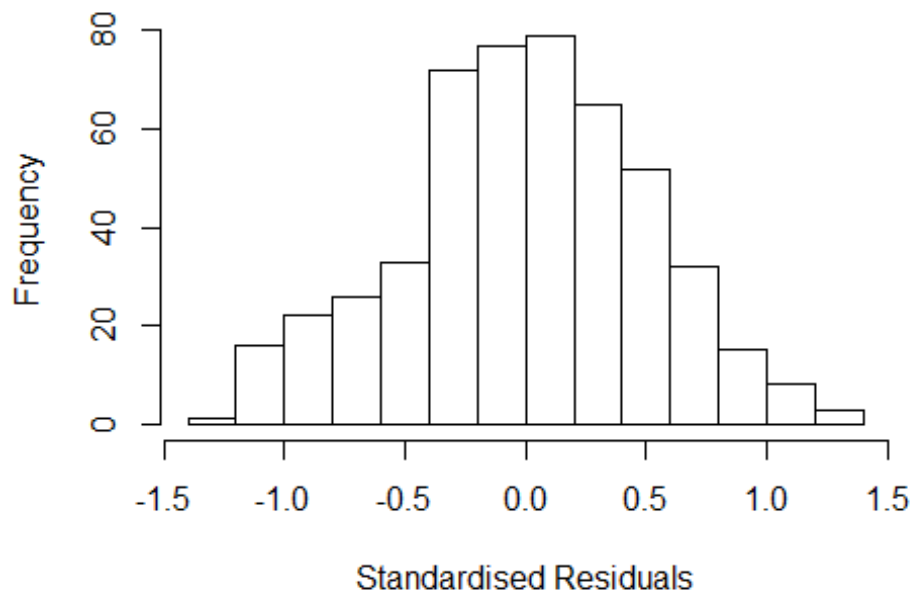
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22975 -0.32389 -0.00483 0.35446 1.40037
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06115 0.14840 7.15 3.2e-12 ***
## FirstAuthorFemale1 0.00898 0.05350 0.17 0.87
## Year1997 -0.19934 0.18659 -1.07 0.29
## Year1998 -0.15262 0.15576 -0.98 0.33
## Year1999 0.07492 0.21427 0.35 0.73
## Year2000 0.05131 0.18791 0.27 0.78
## Year2001 -0.03732 0.17099 -0.22 0.83
## Year2002 -0.19727 0.19169 -1.03 0.30
## Year2003 -0.09338 0.22378 -0.42 0.68
## Year2004 -0.03984 0.19752 -0.20 0.84
## Year2005 0.15953 0.20085 0.79 0.43
## Year2006 -0.12750 0.17656 -0.72 0.47
```

```

## Year2007          0.15962    0.16638    0.96    0.34
## Year2008          -0.08371    0.16905   -0.50    0.62
## Year2009           0.06786    0.17060    0.40    0.69
## Year2010          -0.07559    0.16953   -0.45    0.66
## Year2011           0.08852    0.16144    0.55    0.58
## Year2012          -0.09337    0.15849   -0.59    0.56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.0405, Adjusted R-squared:  0.00676
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 457 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.412  0.864  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      2.00e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## 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



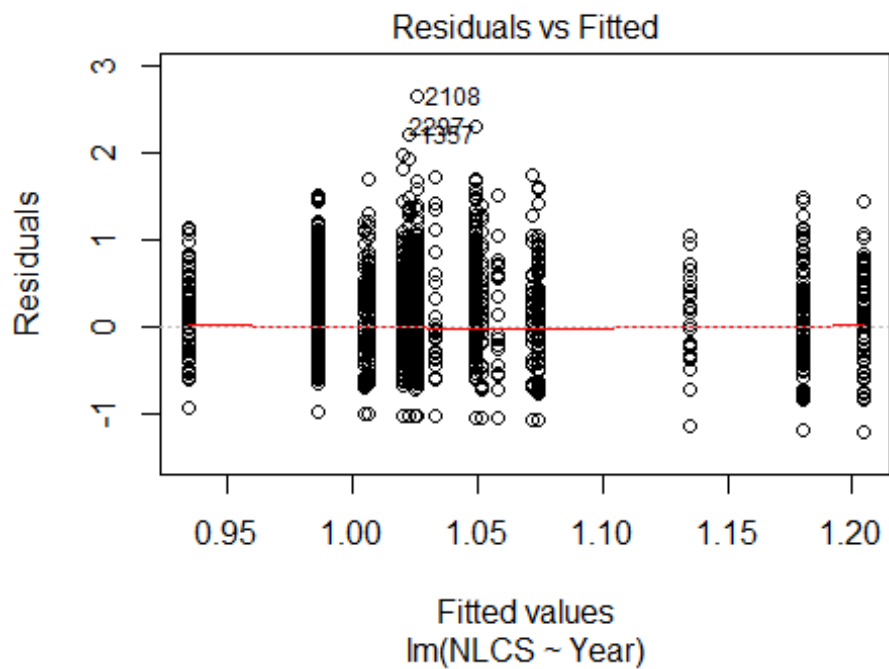
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.25121 -0.31135  0.00316  0.35115  1.38131
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0366     0.1454   7.13 3.7e-12 ***
## LastAuthorFemale1  0.0690     0.0502   1.37  0.17
## Year1997        -0.1983     0.1846  -1.07  0.28
## Year1998        -0.1649     0.1546  -1.07  0.29
## Year1999         0.0773     0.2121   0.36  0.72
## Year2000         0.0380     0.1881   0.20  0.84
## Year2001        -0.0412     0.1692  -0.24  0.81
## Year2002        -0.2050     0.1872  -1.10  0.27
## Year2003        -0.1014     0.2196  -0.46  0.64
## Year2004        -0.0594     0.1950  -0.30  0.76
## Year2005         0.1454     0.1980   0.73  0.46
## Year2006        -0.1439     0.1767  -0.81  0.42
```

```

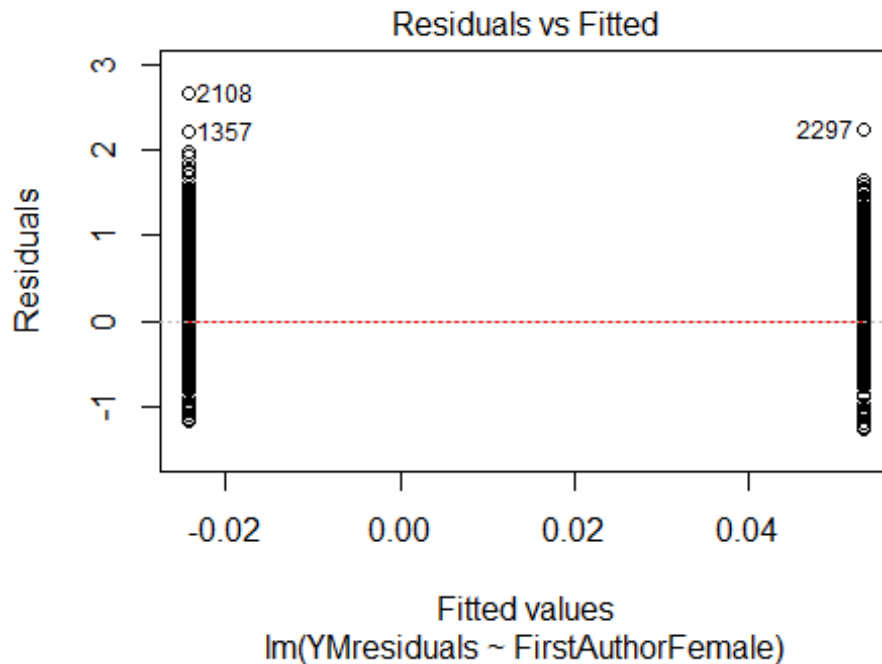
## Year2007          0.1456      0.1646      0.88      0.38
## Year2008          -0.0980     0.1676     -0.58     0.56
## Year2009          0.0531     0.1682     0.32     0.75
## Year2010          -0.0804     0.1693     -0.47     0.64
## Year2011          0.0748     0.1607     0.47     0.64
## Year2012          -0.1077     0.1576     -0.68     0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0448, Adjusted R-squared:  0.0111
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 453 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.418  0.864  0.949  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      2.00e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 501"
## [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
##   40   55   66   39   56   43   85  110   79  109  127  109  222  243  336
## 2011 2012
##  329  381
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32   48   56   29   38   24   70   95   68   95  116   92  178  201  278
## 2011 2012

```

```
## 293 306
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 32 47 56 28 37 24 69 90 65 92 113 90 174 193 268
## 2011 2012
## 287 302
## [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
```

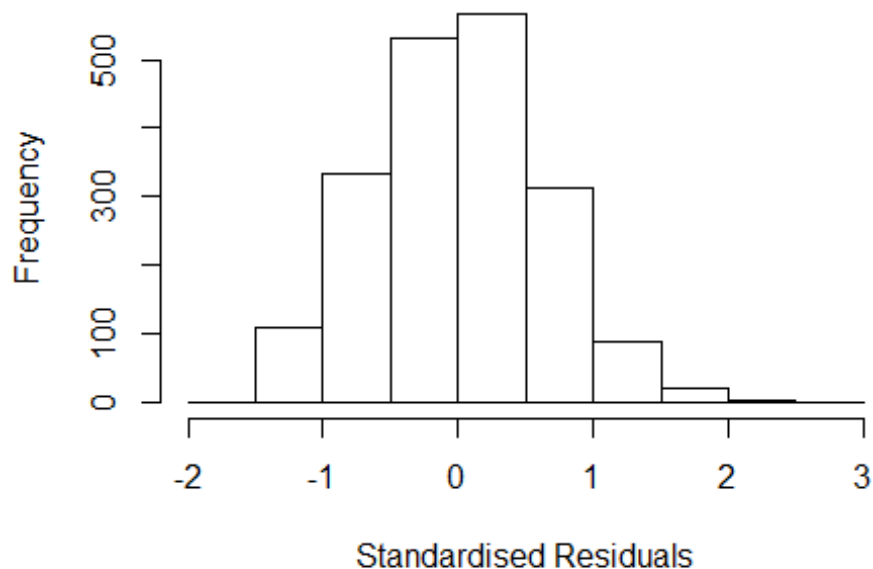


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.25, df = 1, p-value = 0.6
```

```
## [1] "Female first author team size 2018 geometric mean: 1.22004086942605"
## [1] "Male first author team size 2018 geometric mean: 1.22975955393563"
##
## 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: 1.21240417360323"
## [1] "Male last author team size 2018 geometric mean: 1.23393916532574"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9700, 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 3.973 1 1.993
## LastAuthorFemale 3.975 1 1.994
## UniqueAuthors 1.200 4 1.023
## Year 1.219 16 1.006
```

Residuals from first and last author and team size



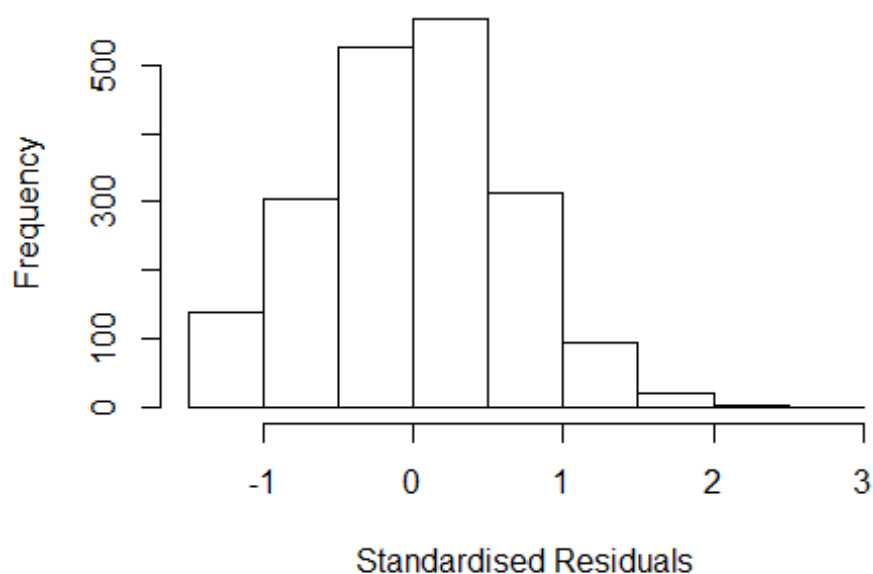
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2108 79955879153 3.665 2011      3312      2      2.555
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5722 -0.4550  0.0112  0.4234  2.5552
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.947288   0.127656   7.42 1.7e-13 ***
## FirstAuthorFemale1 0.073115   0.062298   1.17  0.24068
## LastAuthorFemale1 0.009973   0.062803   0.16  0.87385
## UniqueAuthors2    0.151063   0.043848   3.45 0.00058 ***
## UniqueAuthors3    0.065734   0.099911   0.66  0.51067
## UniqueAuthors4    0.404735   0.185688   2.18  0.02940 *
## UniqueAuthors5   -0.421419   0.214852  -1.96  0.04997 *
## Year1997          0.072159   0.163896   0.44  0.65979
## Year1998          0.066139   0.148594   0.45  0.65630
## Year1999          0.137074   0.176772   0.78  0.43818
```

```

## Year2000      0.048261    0.192152    0.25    0.80172
## Year2001      0.044347    0.211876    0.21    0.83423
## Year2002      0.217390    0.146735    1.48    0.13863
## Year2003     -0.062575    0.142294   -0.44    0.66016
## Year2004      0.001816    0.147510    0.01    0.99018
## Year2005      0.095808    0.141746    0.68    0.49918
## Year2006      0.031921    0.138818    0.23    0.81815
## Year2007      0.158039    0.145329    1.09    0.27697
## Year2008      0.000214    0.135514    0.00    0.99874
## Year2009     -0.011058    0.134872   -0.08    0.93466
## Year2010     -0.024511    0.133034   -0.18    0.85384
## Year2011      0.001439    0.133490    0.01    0.99140
## Year2012      0.034968    0.134835    0.26    0.79540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.0114
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0873 0.8640 0.9520 0.9130 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          5.08e-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 3.979 1          1.995
## LastAuthorFemale 3.975 1          1.994
## Year              1.075 16          1.002

```

Residuals from first and last author



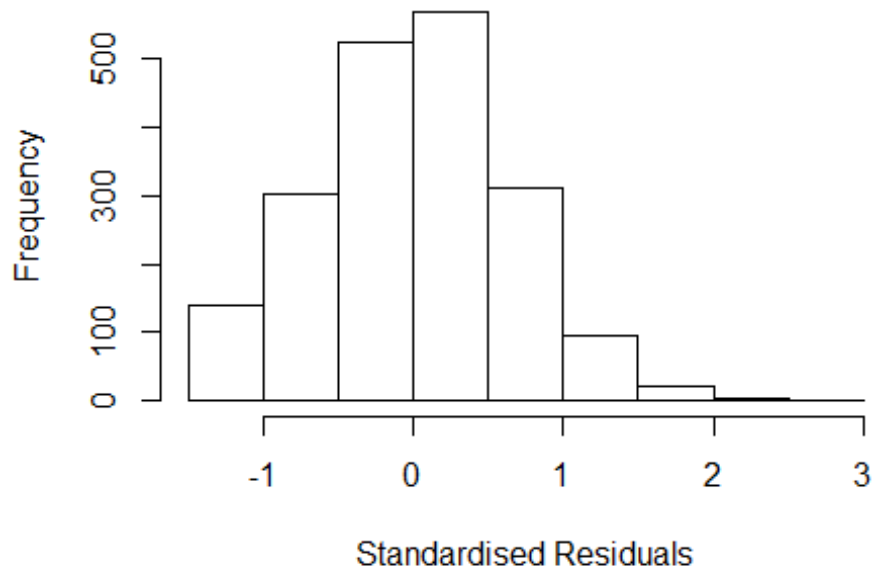
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2108 79955879153 3.665 2011      3312      2      2.698
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27055 -0.44240  0.00818  0.43274  2.69789
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95597    0.12650   7.56 6.3e-14 ***
## FirstAuthorFemale1  0.09258    0.06287   1.47  0.14
## LastAuthorFemale1 -0.00659    0.06346  -0.10  0.92
## Year1997         0.08456    0.16325   0.52  0.60
## Year1998         0.07469    0.14654   0.51  0.61
## Year1999         0.15252    0.17540   0.87  0.38
## Year2000         0.06636    0.19436   0.34  0.73
## Year2001         0.05516    0.21009   0.26  0.79
## Year2002         0.22860    0.14613   1.56  0.12
## Year2003        -0.05829    0.14065  -0.41  0.68
## Year2004         0.00888    0.14708   0.06  0.95
## Year2005         0.10459    0.14087   0.74  0.46
```

```

## Year2006          0.03882    0.13789    0.28    0.78
## Year2007          0.18283    0.14538    1.26    0.21
## Year2008          0.01186    0.13429    0.09    0.93
## Year2009          0.01202    0.13338    0.09    0.93
## Year2010         -0.01540    0.13197   -0.12    0.91
## Year2011          0.01773    0.13236    0.13    0.89
## Year2012          0.05000    0.13378    0.37    0.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.00378
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1820 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0476 0.8630 0.9500 0.9130 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      5.08e-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 1.049 1      1.024
## Year              1.049 16      1.001

```

Residuals from first author

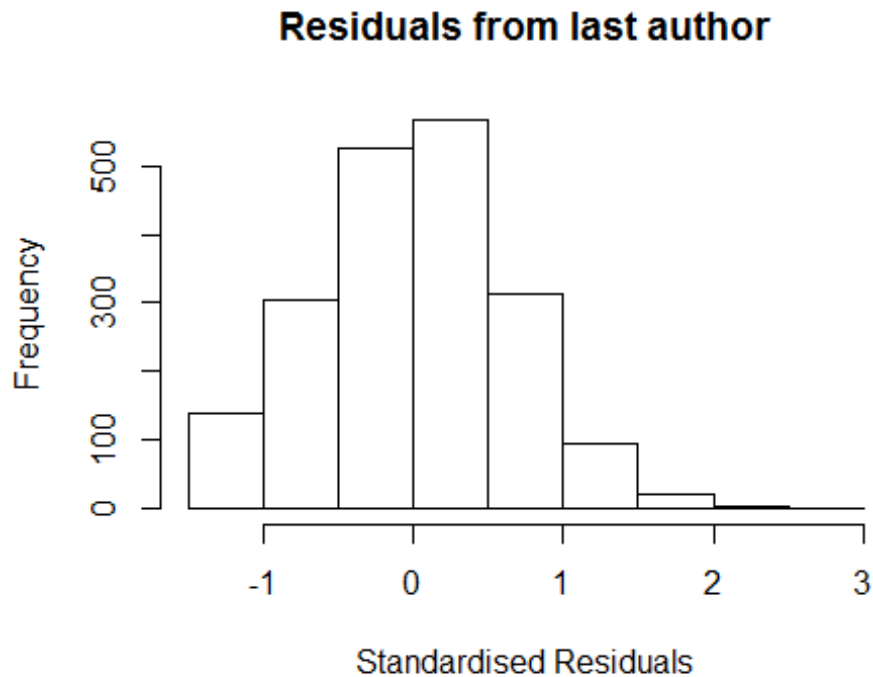


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2108 79955879153 3.665 2011      3312      2      2.698
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27168 -0.44270  0.00756  0.43260  2.69159
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95565    0.12648    7.56 6.4e-14 ***
## FirstAuthorFemale1 0.08697    0.03228    2.69  0.0071 **
## Year1997        0.08493    0.16311    0.52  0.6027
## Year1998        0.07466    0.14656    0.51  0.6105
## Year1999        0.15312    0.17519    0.87  0.3822
## Year2000        0.06640    0.19441    0.34  0.7327
## Year2001        0.05469    0.20998    0.26  0.7945
## Year2002        0.22907    0.14610    1.57  0.1171
## Year2003       -0.05812    0.14067   -0.41  0.6795
## Year2004        0.00904    0.14711    0.06  0.9510
## Year2005        0.10463    0.14086    0.74  0.4577
## Year2006        0.03888    0.13791    0.28  0.7780
```

```

## Year2007          0.18275    0.14540    1.26    0.2089
## Year2008          0.01182    0.13431    0.09    0.9299
## Year2009          0.01205    0.13339    0.09    0.9280
## Year2010         -0.01540    0.13199   -0.12    0.9072
## Year2011          0.01776    0.13237    0.13    0.8933
## Year2012          0.05000    0.13381    0.37    0.7087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0129, Adjusted R-squared:  0.00429
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1822 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0487 0.8630 0.9500 0.9130 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      5.08e-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 1.047 1          1.023
## Year            1.047 16          1.001

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2108 79955879153 3.665 2011      3312      2      2.698
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2657 -0.4482  0.0106  0.4346  2.6129
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9589     0.1263   7.59 4.8e-14 ***
## LastAuthorFemale1  0.0732     0.0326   2.25  0.025 *
## Year1997          0.0886     0.1628   0.54  0.586
## Year1998          0.0706     0.1466   0.48  0.630
## Year1999          0.1606     0.1748   0.92  0.358
## Year2000          0.0674     0.1941   0.35  0.728
## Year2001          0.0534     0.2113   0.25  0.800
## Year2002          0.2335     0.1461   1.60  0.110
## Year2003         -0.0588     0.1406  -0.42  0.676
## Year2004          0.0130     0.1471   0.09  0.930
## Year2005          0.1038     0.1406   0.74  0.460
## Year2006          0.0393     0.1377   0.29  0.775
```

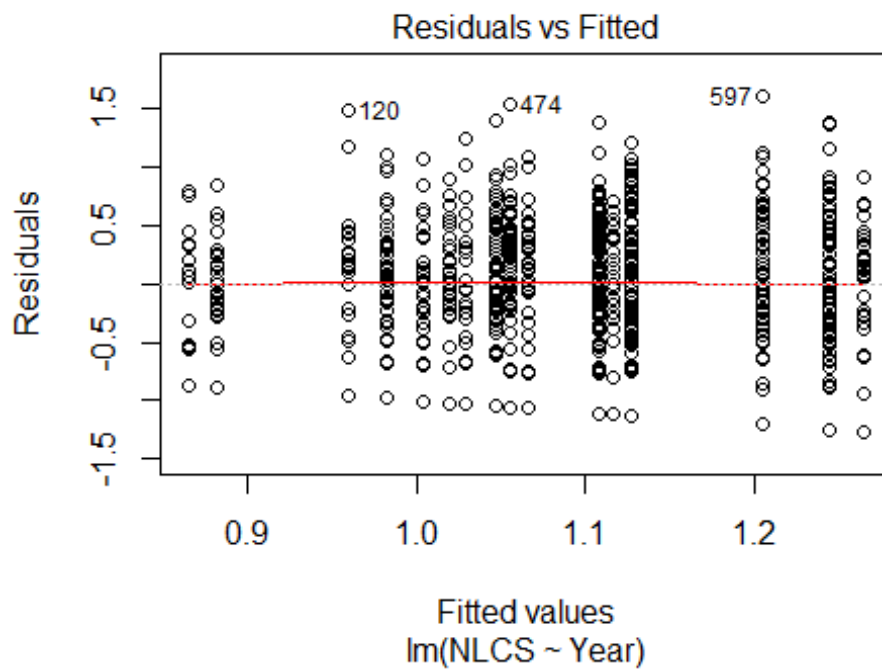


```

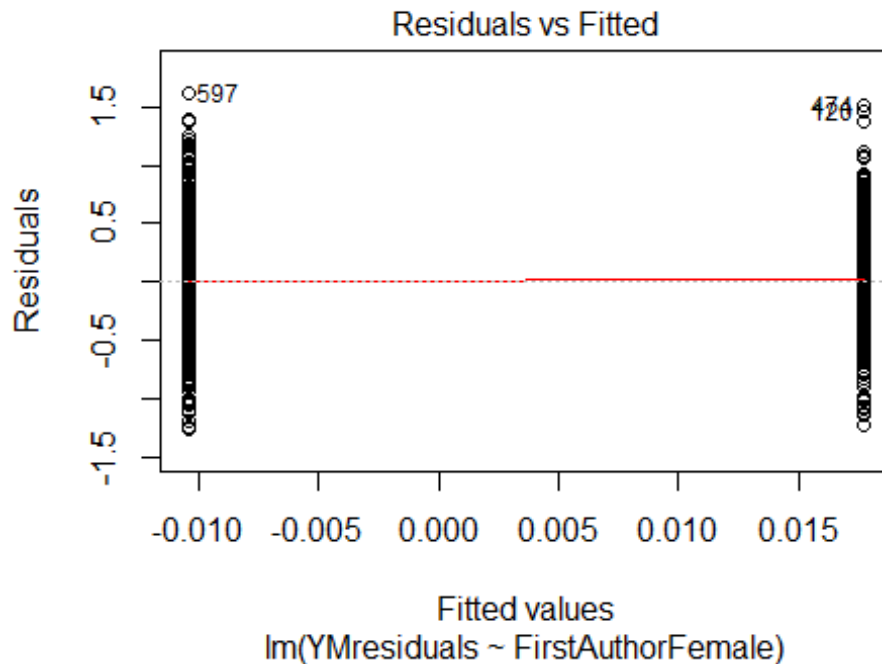
## Year2007          0.1842      0.1453      1.27      0.205
## Year2008          0.0142      0.1342      0.11      0.916
## Year2009          0.0142      0.1332      0.11      0.915
## Year2010         -0.0137      0.1318     -0.10      0.917
## Year2011          0.0200      0.1322      0.15      0.880
## Year2012          0.0511      0.1337      0.38      0.702
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.00312
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1822 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0708 0.8650 0.9510 0.9130 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      5.08e-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: 1967"
## [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
##   27   25   19   32   33   39   51   34   32   58   49   63   94   96  130
## 2011 2012
##  121   92
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   24   17   27   24   28   49   31   26   53   45   58   82   77  113
## 2011 2012

```

```
## 104 82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 24 22 16 27 24 27 48 28 26 52 40 56 74 74 102
## 2011 2012
## 98 77
## [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
```

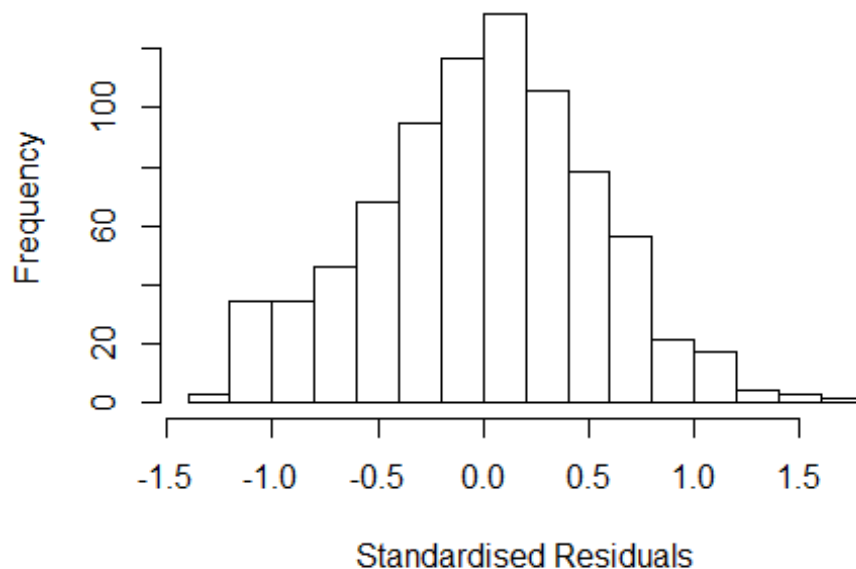


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.84, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 1.94647131340945"
## [1] "Male first author team size 2018 geometric mean: 1.50882423593729"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.72963722156651"
## [1] "Male last author team size 2018 geometric mean: 1.74711461728495"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, 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 1.817 1 1.348
## LastAuthorFemale 1.818 1 1.348
## UniqueAuthors 1.854 4 1.080
## Year 1.893 16 1.020
```

Residuals from first and last author and team size



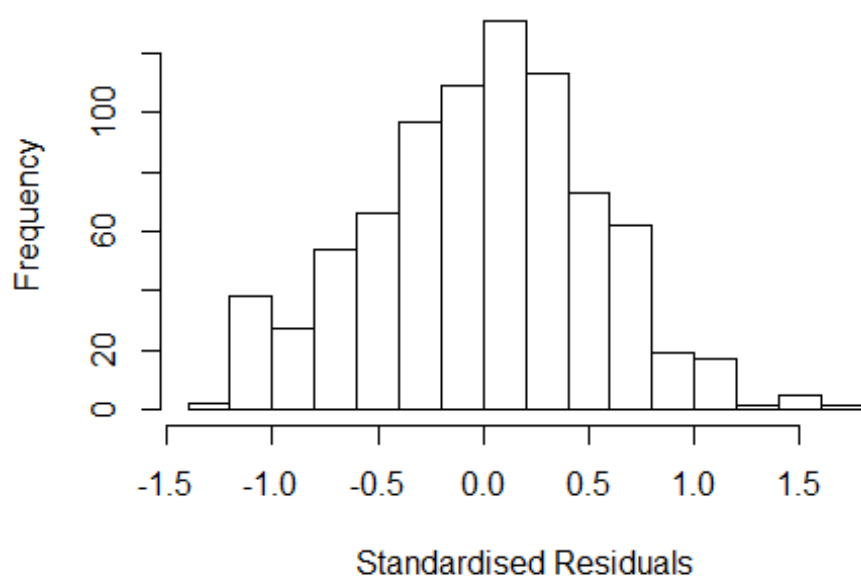
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2771 -0.3496 0.0145 0.3525 1.6390
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95020 0.12488 7.61 7.9e-14 ***
## FirstAuthorFemale1 -0.04056 0.05251 -0.77 0.4401
## LastAuthorFemale1 0.14521 0.05429 2.67 0.0076 **
## UniqueAuthors2 0.03183 0.04600 0.69 0.4892
## UniqueAuthors3 0.10669 0.07026 1.52 0.1293
## UniqueAuthors4 -0.08164 0.14151 -0.58 0.5642
## UniqueAuthors5 -0.15937 0.10857 -1.47 0.1425
## Year1997 0.07635 0.15902 0.48 0.6313
## Year1998 -0.08593 0.17735 -0.48 0.6281
## Year1999 0.01664 0.16801 0.10 0.9211
```

```

## Year2000      -0.03368    0.17541   -0.19    0.8478
## Year2001      -0.10289    0.14558   -0.71    0.4799
## Year2002       0.10077    0.14949    0.67    0.5005
## Year2003       0.30679    0.15711    1.95    0.0512 .
## Year2004       0.14314    0.15814    0.91    0.3657
## Year2005      -0.01043    0.14315   -0.07    0.9419
## Year2006       0.00311    0.14910    0.02    0.9834
## Year2007       0.05692    0.14796    0.38    0.7006
## Year2008       0.21882    0.14242    1.54    0.1248
## Year2009       0.23921    0.14349    1.67    0.0959 .
## Year2010       0.13870    0.13684    1.01    0.3111
## Year2011       0.11552    0.13684    0.84    0.3988
## Year2012       0.04335    0.13796    0.31    0.7535
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0515, Adjusted R-squared:  0.0251
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 740 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.865  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      1.23e-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 1.707 1      1.307
## LastAuthorFemale  1.711 1      1.308
## Year              1.259 16      1.007

```

Residuals from first and last author



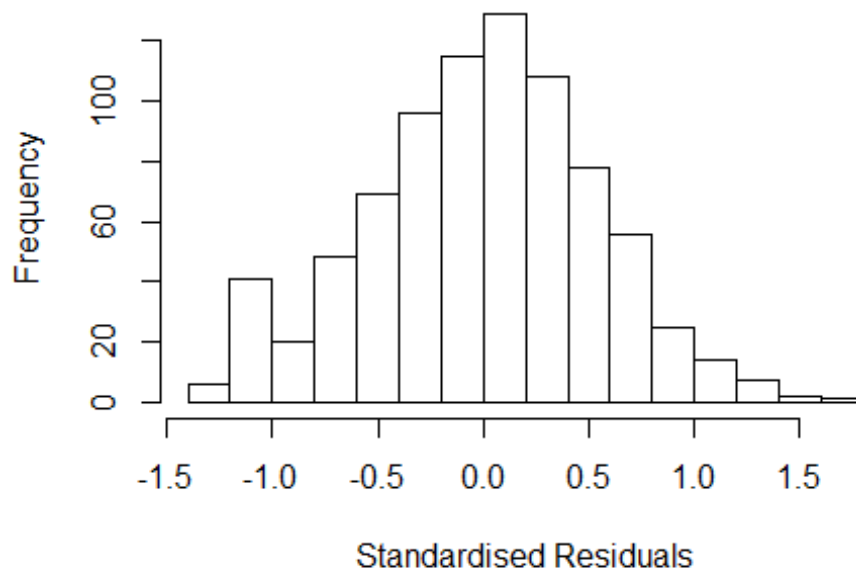
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2742 -0.3510 0.0197 0.3597 1.6212
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.976959 0.122582 7.97 5.5e-15 ***
## FirstAuthorFemale1 -0.035183 0.051751 -0.68 0.497
## LastAuthorFemale1 0.132626 0.053237 2.49 0.013 *
## Year1997 0.073222 0.155771 0.47 0.638
## Year1998 -0.108708 0.174886 -0.62 0.534
## Year1999 -0.000106 0.165634 0.00 0.999
## Year2000 -0.046276 0.172639 -0.27 0.789
## Year2001 -0.107869 0.141339 -0.76 0.446
## Year2002 0.093557 0.147875 0.63 0.527
## Year2003 0.297254 0.153658 1.93 0.053 .
## Year2004 0.131240 0.154817 0.85 0.397
## Year2005 -0.023669 0.140236 -0.17 0.866
```

```

## Year2006          0.014730    0.147085    0.10    0.920
## Year2007          0.045388    0.146336    0.31    0.757
## Year2008          0.209822    0.142368    1.47    0.141
## Year2009          0.220731    0.141425    1.56    0.119
## Year2010          0.106050    0.133211    0.80    0.426
## Year2011          0.103969    0.134242    0.77    0.439
## Year2012          0.035422    0.137857    0.26    0.797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0439, Adjusted R-squared:  0.0223
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.337  0.866  0.952  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      1.23e-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 1.158 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
## -1.2866 -0.3485 0.0112 0.3677 1.6140
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9953 0.1194 8.33 3.4e-16 ***
## FirstAuthorFemale1 0.0367 0.0428 0.86 0.392
## Year1997 0.0827 0.1551 0.53 0.594
## Year1998 -0.1312 0.1742 -0.75 0.452
## Year1999 -0.0121 0.1651 -0.07 0.942
## Year2000 -0.0727 0.1700 -0.43 0.669
## Year2001 -0.1276 0.1387 -0.92 0.358
## Year2002 0.0865 0.1456 0.59 0.552
## Year2003 0.2913 0.1524 1.91 0.056 .
## Year2004 0.1298 0.1507 0.86 0.389
## Year2005 -0.0284 0.1382 -0.21 0.837
## Year2006 0.0198 0.1448 0.14 0.891
```

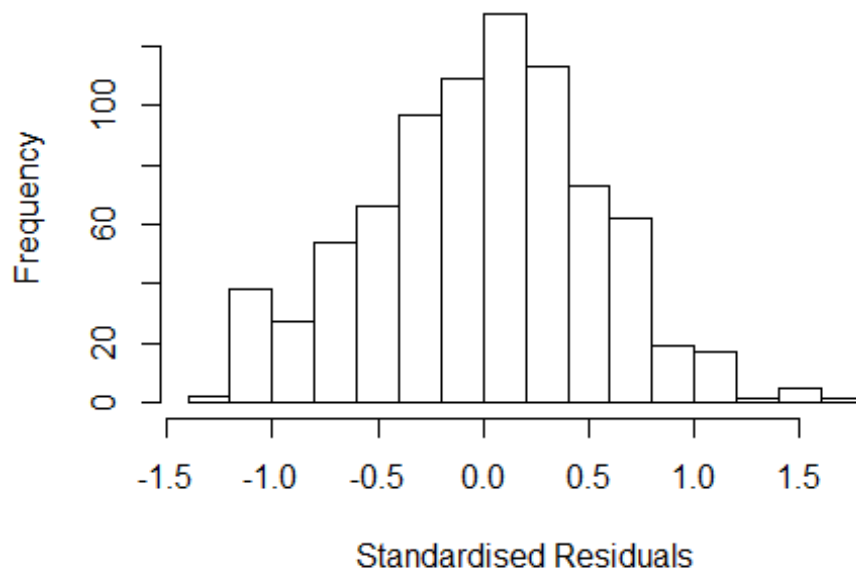


```

## Year2007          0.0418      0.1431      0.29      0.770
## Year2008          0.1987      0.1398      1.42      0.156
## Year2009          0.2373      0.1380      1.72      0.086 .
## Year2010          0.1219      0.1294      0.94      0.346
## Year2011          0.1154      0.1312      0.88      0.379
## Year2012          0.0361      0.1346      0.27      0.789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0359, Adjusted R-squared:  0.0154
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 739 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.341  0.866  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      1.23e-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 1.157 1      1.075
## Year      1.157 16      1.005

```

Residuals from last author



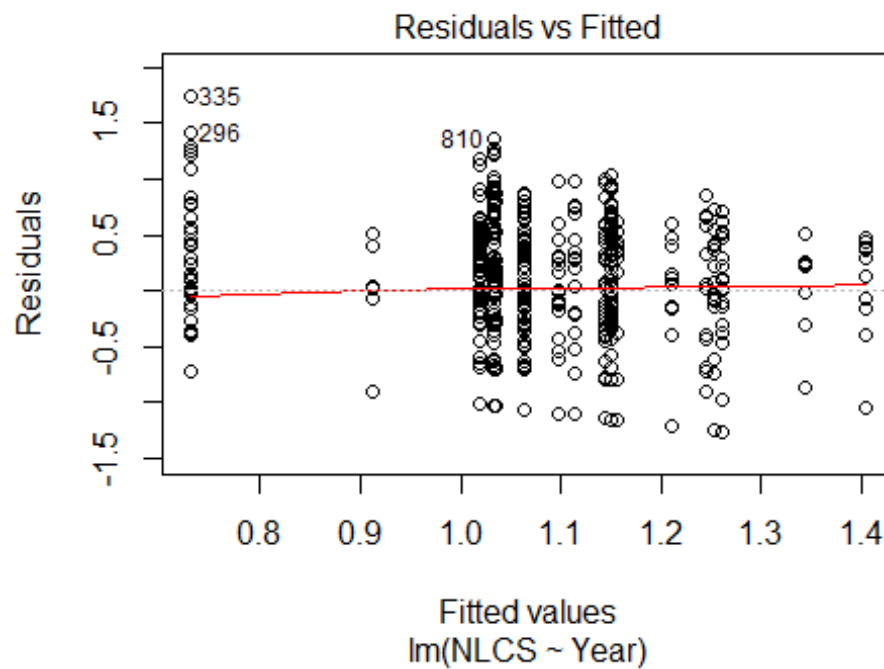
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2657 -0.3509 0.0232 0.3551 1.6303
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97366 0.12260 7.94 6.8e-15 ***
## LastAuthorFemale1 0.11266 0.04384 2.57 0.010 *
## Year1997 0.07588 0.15624 0.49 0.627
## Year1998 -0.10749 0.17576 -0.61 0.541
## Year1999 -0.00315 0.16715 -0.02 0.985
## Year2000 -0.05063 0.17358 -0.29 0.771
## Year2001 -0.11068 0.14180 -0.78 0.435
## Year2002 0.09227 0.14852 0.62 0.535
## Year2003 0.29200 0.15387 1.90 0.058 .
## Year2004 0.12918 0.15543 0.83 0.406
## Year2005 -0.02457 0.14091 -0.17 0.862
## Year2006 0.01043 0.14764 0.07 0.944
```

```

## Year2007      0.04046    0.14676    0.28    0.783
## Year2008      0.20407    0.14327    1.42    0.155
## Year2009      0.21789    0.14173    1.54    0.125
## Year2010      0.10745    0.13346    0.81    0.421
## Year2011      0.09892    0.13491    0.73    0.464
## Year2012      0.03177    0.13847    0.23    0.819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0432, Adjusted R-squared:  0.0228
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 741 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.331  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      1.23e-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: 815"
## [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
##   13   15   17   15   30   19   19   14   23   26   36   24   89  100   89
## 2011 2012
##  102  122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   12   11   12   15    8   18   12   20   21   30   21   79   91   83
## 2011 2012

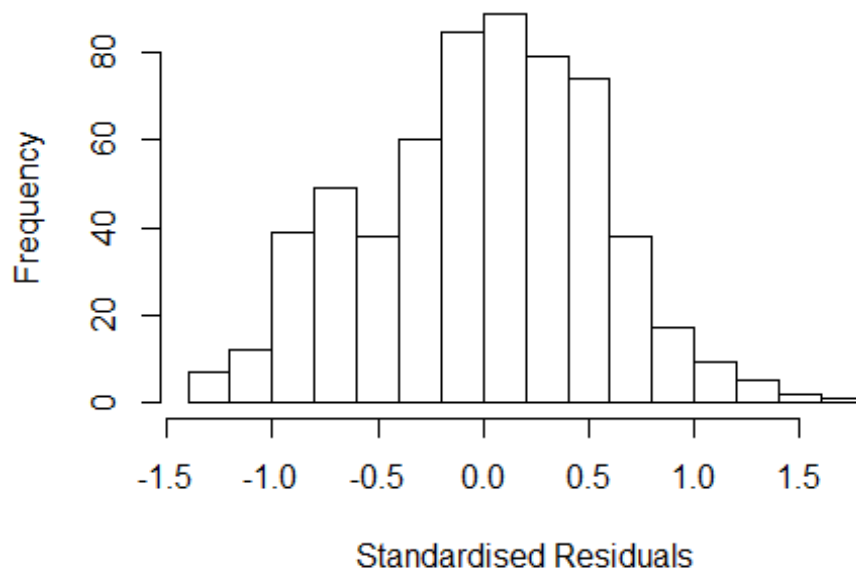
```

```
##      83  106
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   12   11   12   15    8   18   12   19   20   29   19   78   86   80
## 2011 2012
##    77  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 = 1.8, df = 1, p-value = 0.2
```


Residuals from first and last author and team size



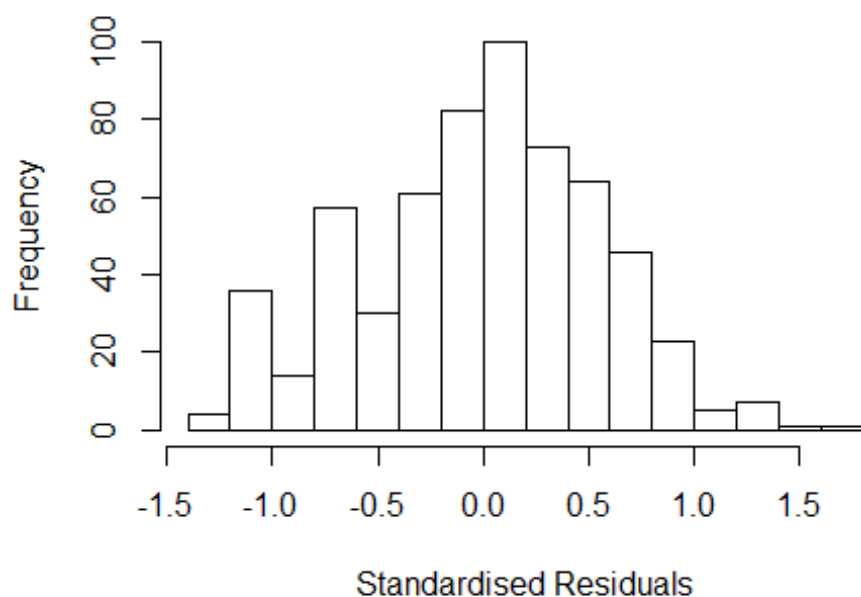
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2992 -0.3783 0.0251 0.3823 1.6105
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3118 0.1774 7.39 5e-13 ***
## FirstAuthorFemale1 0.0700 0.0607 1.15 0.24941
## LastAuthorFemale1 -0.0905 0.0616 -1.47 0.14263
## UniqueAuthors2 0.1932 0.0610 3.17 0.00161 **
## UniqueAuthors3 0.2063 0.0767 2.69 0.00736 **
## UniqueAuthors4 -0.0597 0.1099 -0.54 0.58746
## UniqueAuthors5 -0.0840 0.0858 -0.98 0.32814
## Year1997 -0.1092 0.2526 -0.43 0.66577
## Year1998 0.0159 0.2277 0.07 0.94443
## Year1999 -0.0809 0.2123 -0.38 0.70324
```

```

## Year2000          -0.0800      0.2139   -0.37   0.70857
## Year2001          -0.4437      0.2125   -2.09   0.03726 *
## Year2002          -0.2215      0.2122   -1.04   0.29713
## Year2003          -0.0849      0.2780   -0.31   0.76018
## Year2004          -0.1312      0.2083   -0.63   0.52896
## Year2005          -0.2928      0.2114   -1.38   0.16666
## Year2006          -0.2537      0.2045   -1.24   0.21525
## Year2007          -0.2128      0.2394   -0.89   0.37459
## Year2008          -0.6575      0.1912   -3.44   0.00063 ***
## Year2009          -0.3438      0.1855   -1.85   0.06425 .
## Year2010          -0.1853      0.1868   -0.99   0.32164
## Year2011          -0.3197      0.1854   -1.72   0.08514 .
## Year2012          -0.3695      0.1922   -1.92   0.05497 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.554
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0803
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 546 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.378  0.877  0.943  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          1.66e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.927 1          1.388
## LastAuthorFemale 1.706 1          1.306
## Year              1.388 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
## -1.3015 -0.3671 0.0225 0.3779 1.7711
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3743 0.1636 8.40 3.4e-16 ***
## FirstAuthorFemale1 0.0727 0.0633 1.15 0.25168
## LastAuthorFemale1 -0.1120 0.0634 -1.77 0.07795 .
## Year1997 -0.1457 0.2446 -0.60 0.55163
## Year1998 0.0536 0.2140 0.25 0.80241
## Year1999 -0.1239 0.2012 -0.62 0.53839
## Year2000 -0.0729 0.2088 -0.35 0.72726
## Year2001 -0.4480 0.2055 -2.18 0.02968 *
## Year2002 -0.2608 0.1976 -1.32 0.18729
## Year2003 -0.0988 0.2628 -0.38 0.70725
## Year2004 -0.1094 0.1973 -0.55 0.57946
## Year2005 -0.2902 0.2005 -1.45 0.14843
```

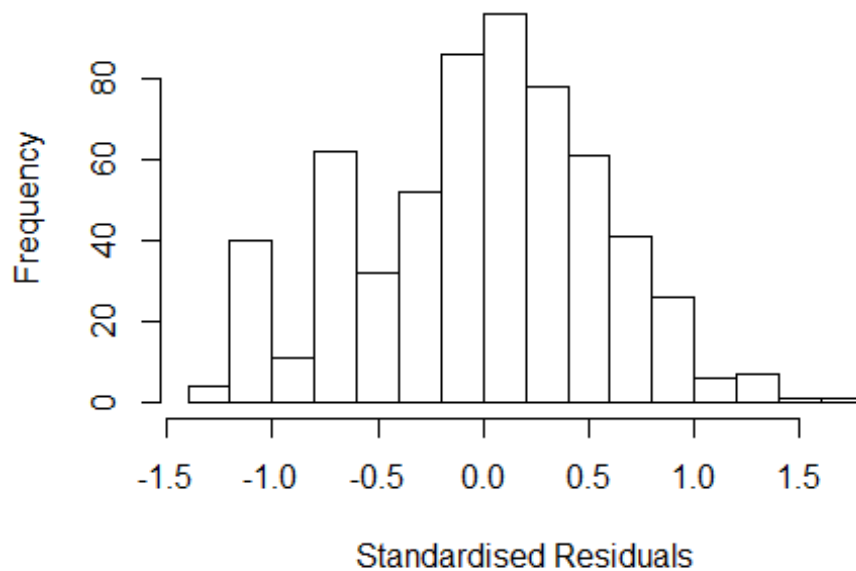


```

## Year2006          -0.2403      0.1934    -1.24    0.21451
## Year2007          -0.1907      0.2320    -0.82    0.41146
## Year2008          -0.6745      0.1813    -3.72    0.00022 ***
## Year2009          -0.3337      0.1733    -1.93    0.05464 .
## Year2010          -0.1854      0.1748    -1.06    0.28937
## Year2011          -0.3086      0.1735    -1.78    0.07582 .
## Year2012          -0.3362      0.1808    -1.86    0.06346 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0873, Adjusted R-squared:  0.0593
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 549 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.300  0.864  0.948  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      1.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.322 1      1.150
## Year              1.322 16      1.009

```

Residuals from first author



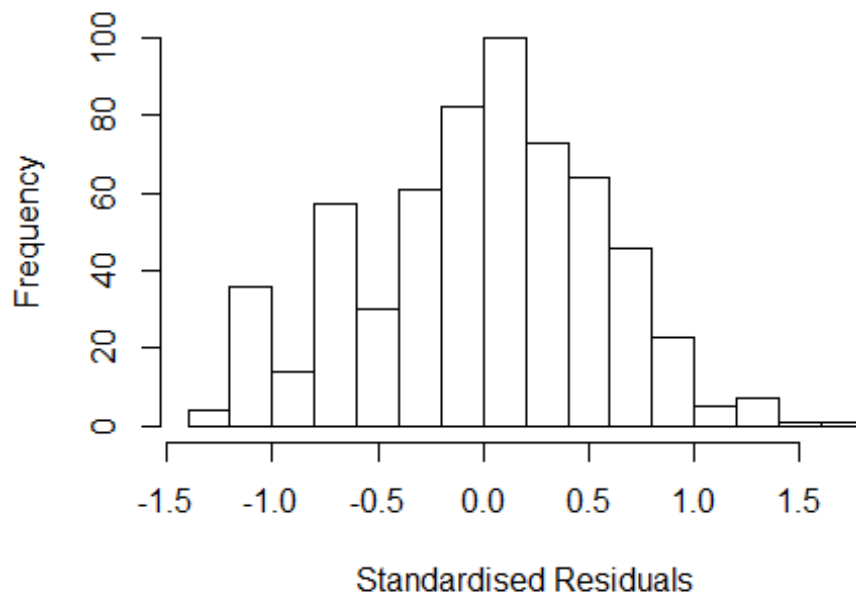
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2907 -0.3783 0.0296 0.3703 1.7882
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36689 0.16173 8.45 2.3e-16 ***
## FirstAuthorFemale1 0.00623 0.05268 0.12 0.90594
## Year1997 -0.15260 0.24525 -0.62 0.53404
## Year1998 0.06155 0.21201 0.29 0.77166
## Year1999 -0.12041 0.19883 -0.61 0.54504
## Year2000 -0.07621 0.20600 -0.37 0.71154
## Year2001 -0.43520 0.20822 -2.09 0.03704 *
## Year2002 -0.26302 0.19514 -1.35 0.17822
## Year2003 -0.11940 0.26032 -0.46 0.64666
## Year2004 -0.10016 0.19552 -0.51 0.60864
## Year2005 -0.29176 0.19931 -1.46 0.14378
## Year2006 -0.24170 0.19184 -1.26 0.20822
```

```

## Year2007          -0.18859    0.22982   -0.82  0.41222
## Year2008          -0.68404    0.17967   -3.81  0.00016 ***
## Year2009          -0.33913    0.17197   -1.97  0.04908 *
## Year2010          -0.20047    0.17314   -1.16  0.24741
## Year2011          -0.31880    0.17125   -1.86  0.06315 .
## Year2012          -0.33728    0.17927   -1.88  0.06041 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0823, Adjusted R-squared:  0.0557
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 546 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.293  0.870  0.946  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      1.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.177 1          1.085
## Year            1.177 16          1.005

```

Residuals from last author



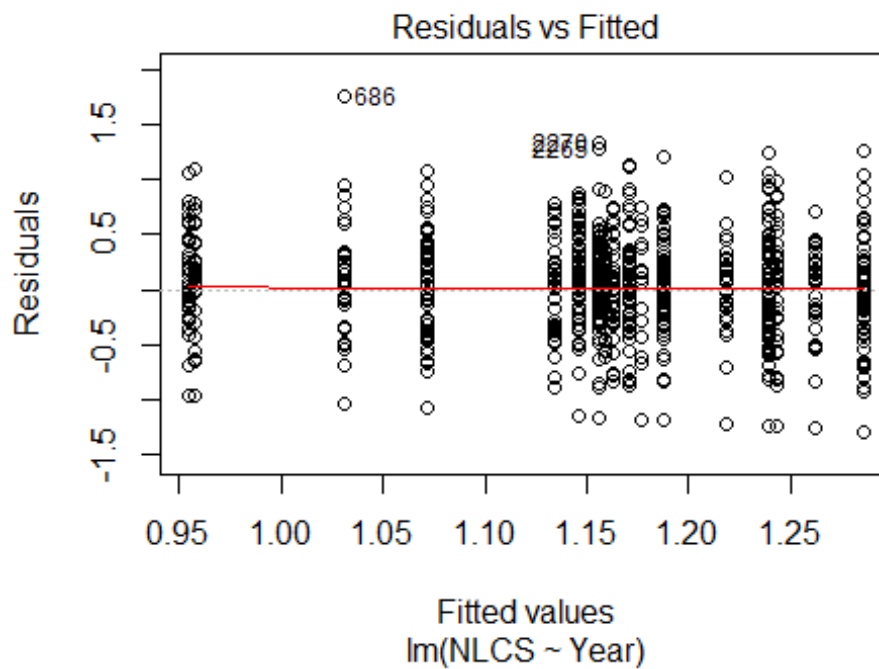
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3073 -0.3849 0.0232 0.3767 1.7595
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3784 0.1648 8.36 4.5e-16 ***
## LastAuthorFemale1 -0.0676 0.0528 -1.28 0.20130
## Year1997 -0.1421 0.2443 -0.58 0.56120
## Year1998 0.0728 0.2122 0.34 0.73178
## Year1999 -0.1262 0.2029 -0.62 0.53430
## Year2000 -0.0711 0.2107 -0.34 0.73598
## Year2001 -0.4423 0.2084 -2.12 0.03423 *
## Year2002 -0.2515 0.1993 -1.26 0.20734
## Year2003 -0.0987 0.2631 -0.38 0.70767
## Year2004 -0.1087 0.1988 -0.55 0.58451
## Year2005 -0.2890 0.2015 -1.43 0.15190
## Year2006 -0.2384 0.1946 -1.23 0.22095
```

```

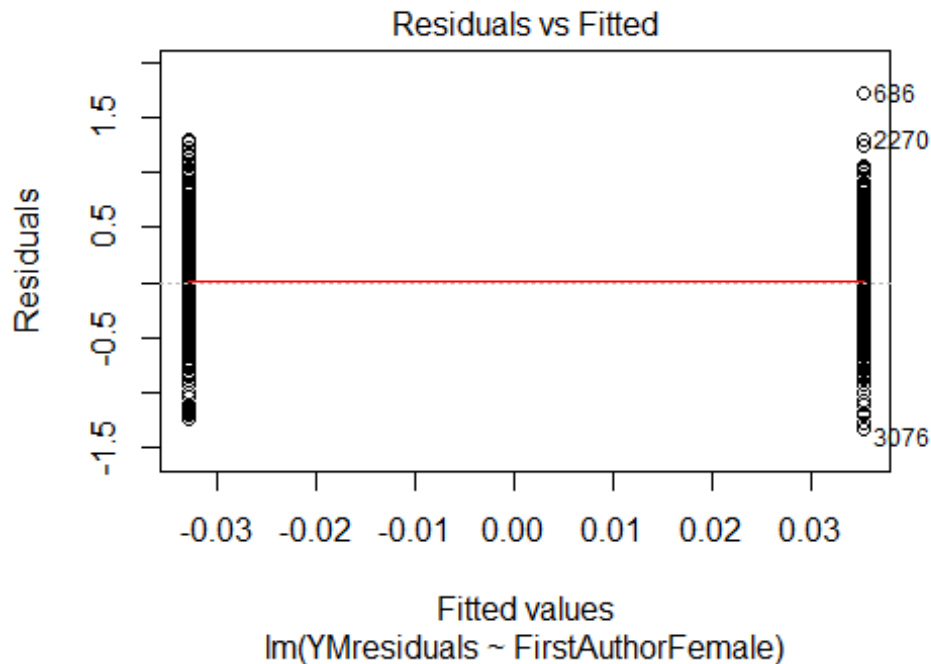
## Year2007          -0.1751      0.2324   -0.75   0.45159
## Year2008          -0.6669      0.1823   -3.66   0.00028 ***
## Year2009          -0.3239      0.1743   -1.86   0.06358 .
## Year2010          -0.1850      0.1761   -1.05   0.29384
## Year2011          -0.2973      0.1746   -1.70   0.08905 .
## Year2012          -0.3225      0.1810   -1.78   0.07532 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.566
## Multiple R-squared:  0.0846, Adjusted R-squared:  0.0581
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 543 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.313  0.864  0.947  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      1.66e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 604"
## [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
##   203  198  162  163  173  156  164  165  157  118  203  216  169  179  186
## 2011 2012
##   239  213
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   29   16   37   33   36   51   39   35   39   65   85   67   72   76
## 2011 2012

```

```
## 91 93
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 27 14 31 25 32 42 35 32 35 58 77 62 58 67
## 2011 2012
## 79 81
## [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
```

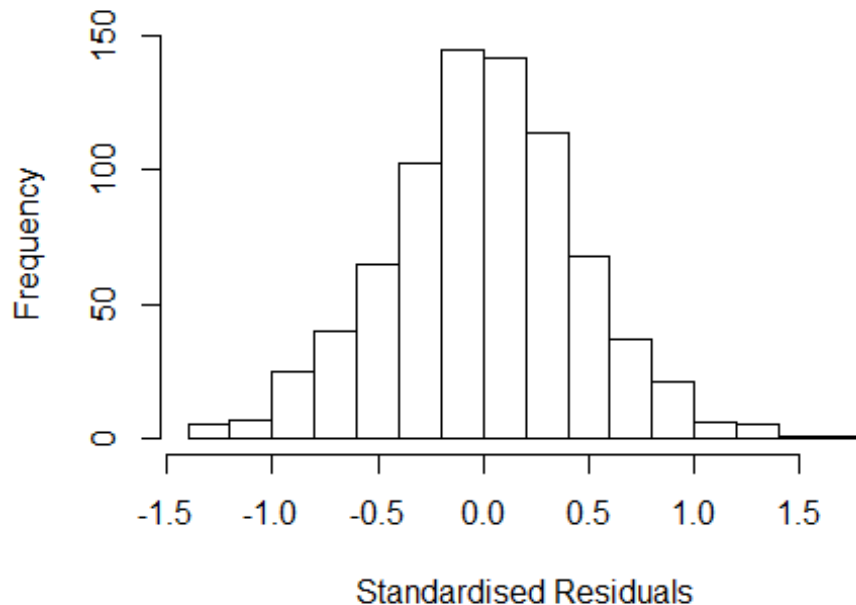


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.4, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 4.04747686529912"
## [1] "Male first author team size 2018 geometric mean: 4.14212297178353"
##
## 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] "Female last author team size 2018 geometric mean: 3.78099651368626"
## [1] "Male last author team size 2018 geometric mean: 4.29350764443712"
##
## 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 1.172 1          1.083
## LastAuthorFemale  1.154 1          1.074
## UniqueAuthors     1.568 4          1.058
## Year               1.792 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
## -1.38276 -0.29995  0.00142  0.29262  1.67149
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0044    0.1229   8.17 1.2e-15 ***
## FirstAuthorFemale1  0.0441    0.0347   1.27  0.20379
## LastAuthorFemale1  0.0753    0.0408   1.84  0.06562 .
## UniqueAuthors2    0.2453    0.0839   2.92  0.00356 **
## UniqueAuthors3    0.2637    0.0719   3.67  0.00026 ***
## UniqueAuthors4    0.2893    0.0759   3.81  0.00015 ***
## UniqueAuthors5    0.2535    0.0682   3.72  0.00021 ***
## Year1997          0.0578    0.1401   0.41  0.68005
## Year1998         -0.0586    0.1675  -0.35  0.72642
## Year1999         -0.2556    0.1443  -1.77  0.07683 .
```

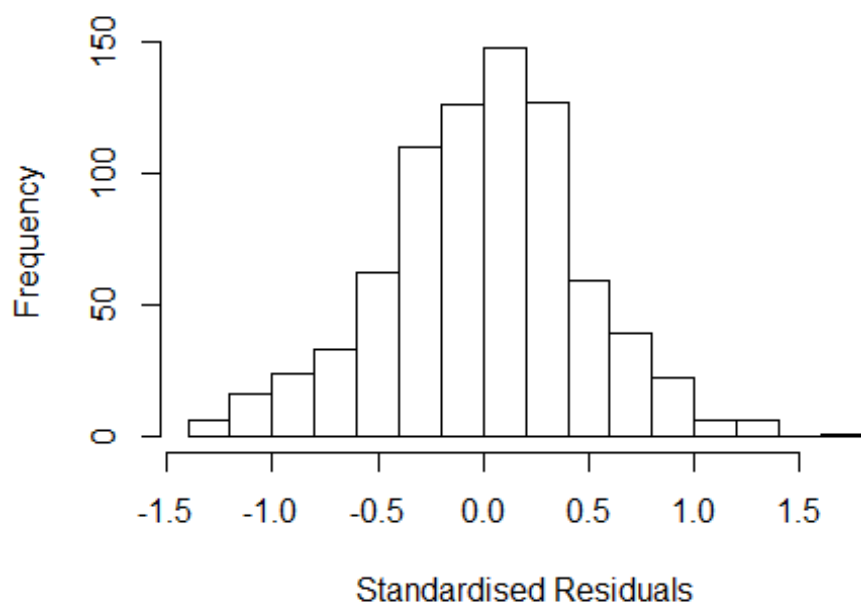


```

## Year2000          -0.1108      0.1398   -0.79   0.42829
## Year2001           0.0436      0.1331    0.33   0.74339
## Year2002          -0.2974      0.1361   -2.18   0.02923 *
## Year2003          -0.3201      0.1362   -2.35   0.01901 *
## Year2004          -0.1792      0.1390   -1.29   0.19787
## Year2005          -0.1460      0.1291   -1.13   0.25815
## Year2006          -0.2357      0.1313   -1.80   0.07296 .
## Year2007          -0.1100      0.1196   -0.92   0.35829
## Year2008          -0.1609      0.1251   -1.29   0.19867
## Year2009          -0.0705      0.1226   -0.57   0.56559
## Year2010          -0.0385      0.1272   -0.30   0.76239
## Year2011          -0.1339      0.1196   -1.12   0.26324
## Year2012          -0.0730      0.1236   -0.59   0.55476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0878, Adjusted R-squared:  0.0615
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 719 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.117  0.869  0.951  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          1.27e-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 1.132 1          1.064
## LastAuthorFemale  1.080 1          1.039
## Year              1.222 16          1.006

```

Residuals from first and last author



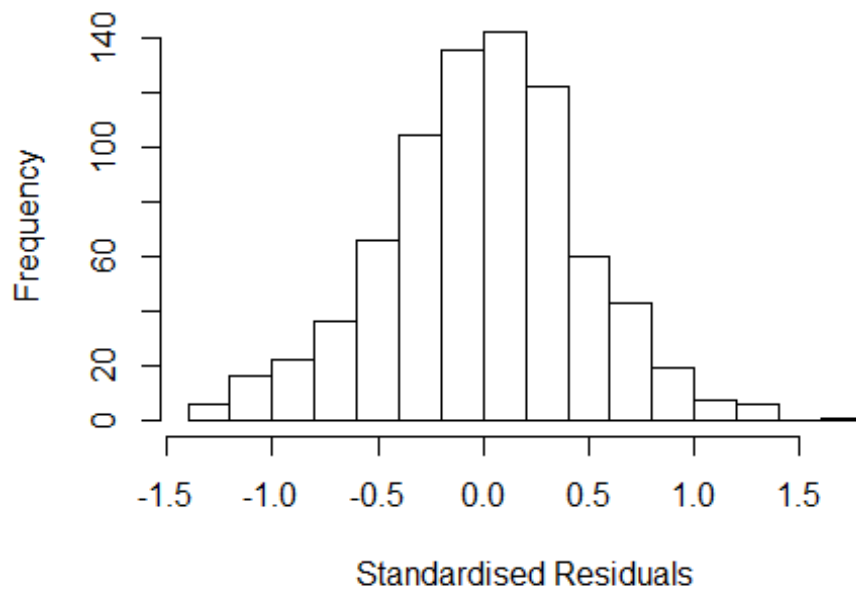
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3302 -0.2993  0.0112  0.2873  1.7050
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19744    0.11068   10.82  <2e-16 ***
## FirstAuthorFemale1  0.06675    0.03484    1.92   0.056 .
## LastAuthorFemale1  0.05998    0.04042    1.48   0.138
## Year1997         0.05177    0.13860    0.37   0.709
## Year1998        -0.05603    0.17469   -0.32   0.748
## Year1999        -0.24417    0.14622   -1.67   0.095 .
## Year2000        -0.06659    0.13804   -0.48   0.630
## Year2001         0.06109    0.12976    0.47   0.638
## Year2002        -0.26462    0.13678   -1.93   0.053 .
## Year2003        -0.34090    0.13707   -2.49   0.013 *
## Year2004        -0.12307    0.13871   -0.89   0.375
## Year2005        -0.09909    0.12628   -0.78   0.433
```

```

## Year2006      -0.19302    0.12941   -1.49    0.136
## Year2007      -0.08146    0.11847   -0.69    0.492
## Year2008      -0.11308    0.12551   -0.90    0.368
## Year2009      -0.04511    0.12304   -0.37    0.714
## Year2010       0.00603    0.12753    0.05    0.962
## Year2011      -0.10443    0.12048   -0.87    0.386
## Year2012      -0.02550    0.12272   -0.21    0.835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0528, Adjusted R-squared:  0.0305
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 719 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0979 0.8650 0.9540 0.8960 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      1.27e-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 1.14 1      1.068
## Year              1.14 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
## -1.28694 -0.29443 0.00944 0.28470 1.75046
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20871 0.11154 10.84 <2e-16 ***
## FirstAuthorFemale1 0.07019 0.03515 2.00 0.046 *
## Year1997 0.04650 0.14023 0.33 0.740
## Year1998 -0.03954 0.17452 -0.23 0.821
## Year1999 -0.24436 0.14661 -1.67 0.096 .
## Year2000 -0.07133 0.13836 -0.52 0.606
## Year2001 0.05854 0.13004 0.45 0.653
## Year2002 -0.26067 0.13820 -1.89 0.060 .
## Year2003 -0.33528 0.13789 -2.43 0.015 *
## Year2004 -0.11663 0.14145 -0.82 0.410
## Year2005 -0.09960 0.12739 -0.78 0.435
## Year2006 -0.19103 0.13082 -1.46 0.145
```

```

## Year2007          -0.08287    0.11966   -0.69    0.489
## Year2008          -0.11028    0.12587   -0.88    0.381
## Year2009          -0.04438    0.12385   -0.36    0.720
## Year2010           0.00804    0.12873    0.06    0.950
## Year2011          -0.09938    0.12127   -0.82    0.413
## Year2012          -0.02434    0.12406   -0.20    0.844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0492, Adjusted R-squared:  0.0281
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.075  0.862  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      1.27e-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 1.09 1      1.044
## Year              1.09 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
## -1.3402 -0.2862  0.0168  0.2768  1.7420

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21801    0.11047   11.03  <2e-16 ***
## LastAuthorFemale1 0.06495    0.04038    1.61   0.108
## Year1997        0.05728    0.13874    0.41   0.680
## Year1998       -0.03820    0.17779   -0.21   0.830
## Year1999       -0.23998    0.14640   -1.64   0.102
## Year2000       -0.06013    0.13671   -0.44   0.660
## Year2001        0.05678    0.13084    0.43   0.664
## Year2002       -0.26657    0.13645   -1.95   0.051 .
## Year2003       -0.32978    0.13721   -2.40   0.016 *
## Year2004       -0.11478    0.13858   -0.83   0.408
## Year2005       -0.09398    0.12711   -0.74   0.460
## Year2006       -0.18229    0.13031   -1.40   0.162
## Year2007       -0.06669    0.11904   -0.56   0.575
## Year2008       -0.10410    0.12585   -0.83   0.408
## Year2009       -0.03185    0.12301   -0.26   0.796
## Year2010        0.02877    0.12675    0.23   0.820
## Year2011       -0.09204    0.12101   -0.76   0.447
## Year2012       -0.00767    0.12324   -0.06   0.950
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0483, Adjusted R-squared:  0.0272
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0784 0.8650 0.9550 0.8960 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           1.27e-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: 785"
## [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
##    7    5    7    2    5   12    3    9    7    4    3    5    9    8    5
## 2011 2012
##    6    6
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    0    4    0    0    3    0    2    0    0    0    2    3    2    2
## 2011 2012
##    2    4
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    0    4    0    0    2    0    2    0    0    0    1    3    2    2
## 2011 2012
##    2    3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.82523450049477"
## [1] "Male first 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 = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3"
## [1] "Male last author team size 2018 geometric mean: 2.70659989709772"

## 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 = 8.5, 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"

```

```

## 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"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 22"
## [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
## 15 3 7 7 5 4 4 3 8 11 19 7 7 19 8
## 2011 2012
## 9 13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 0 1 2 0 0 0 1 1 6 3 3 1 2 1
## 2011 2012
## 0 1
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 0 0 1 0 0 0 1 1 6 3 3 1 2 1
## 2011 2012
## 0 1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5.47722557505166"
## [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 = 6, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.47722557505166"
## [1] "Male last 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 = 6, 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"
## [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: 22"
## [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
## 34 25 22 17 15 12 16 18 8 21 24 25 24 39 48
## 2011 2012
## 38 45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 3 6 2 1 2 2 3 2 8 7 16 8 9 16
## 2011 2012
## 20 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 3 5 2 1 1 1 3 2 8 7 15 7 7 16
## 2011 2012
## 17 19
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.92181043974187"
## [1] "Male first author team size 2018 geometric mean: 2.49546486517789"

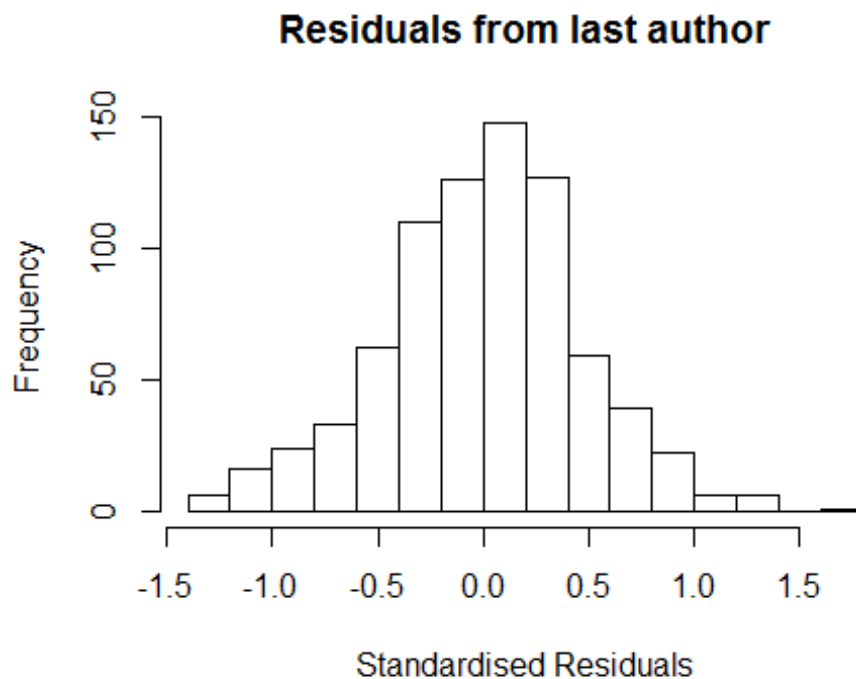
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##

```

```
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 230, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.29349056159138"
## [1] "Male last author team size 2018 geometric mean: 3.44053551695782"

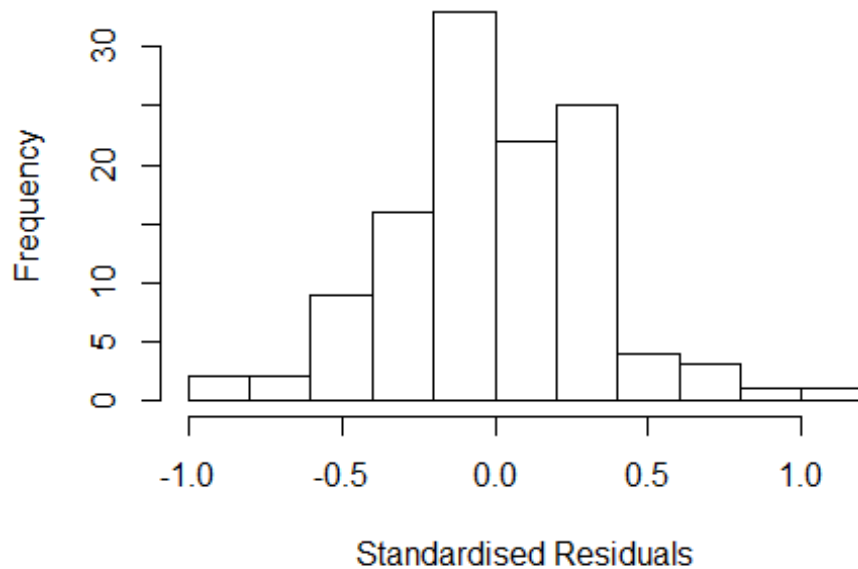
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.02
## 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	2.802e-01	1	0.5293
LastAuthorFemale	-6.488e+13	1	NaN
UniqueAuthors	-6.122e+29	4	NaN
Year	-2.283e+30	16	NaN

Residuals from first and last author and team size



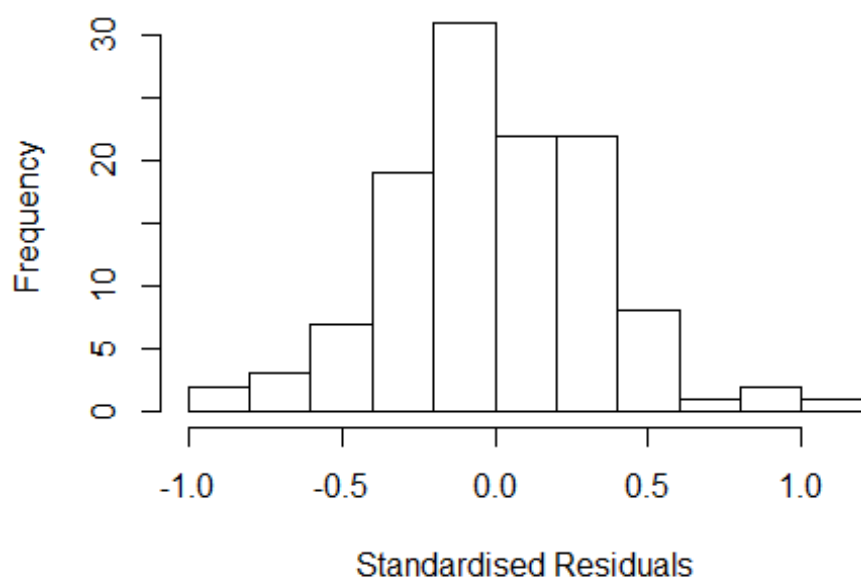
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.888552 -0.197932 -0.000563 0.238451 1.073564
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16079 0.10971 10.58 <2e-16 ***
## FirstAuthorFemale1 0.17561 0.07581 2.32 0.0227 *
## LastAuthorFemale1 0.17851 0.08381 2.13 0.0358 *
## UniqueAuthors2 -0.11574 0.16048 -0.72 0.4725
## UniqueAuthors3 -0.20788 0.14924 -1.39 0.1669
## UniqueAuthors4 -0.16742 0.15633 -1.07 0.2869
## UniqueAuthors5 -0.03347 0.17078 -0.20 0.8450
## Year1997 0.38860 0.15526 2.50 0.0140 *
## Year1998 -0.01661 0.15904 -0.10 0.9170
## Year1999 0.13622 0.27280 0.50 0.6187
```

```

## Year2000      0.16358      0.07352      2.22      0.0285 *
## Year2001      0.14821      0.10971      1.35      0.1799
## Year2002     -0.26331      0.09394     -2.80      0.0061 **
## Year2003     -0.01904      0.11247     -0.17      0.8660
## Year2004      0.21821      0.13231      1.65      0.1024
## Year2005      0.11865      0.11263      1.05      0.2948
## Year2006      0.10222      0.17547      0.58      0.5616
## Year2007      0.06039      0.09379      0.64      0.5212
## Year2008     -0.07902      0.11239     -0.70      0.4837
## Year2009     -0.00619      0.12906     -0.05      0.9618
## Year2010     -0.06763      0.10052     -0.67      0.5027
## Year2011     -0.14630      0.14728     -0.99      0.3231
## Year2012     -0.06435      0.11200     -0.57      0.5669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.187, Adjusted R-squared:  -0.00176
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.349  0.906  0.948  0.916  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      8.47e-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 2.642e+00 1      1.626
## LastAuthorFemale  3.647e+00 1      1.910
## Year              1.015e+14 16      2.740

```

Residuals from first and last author



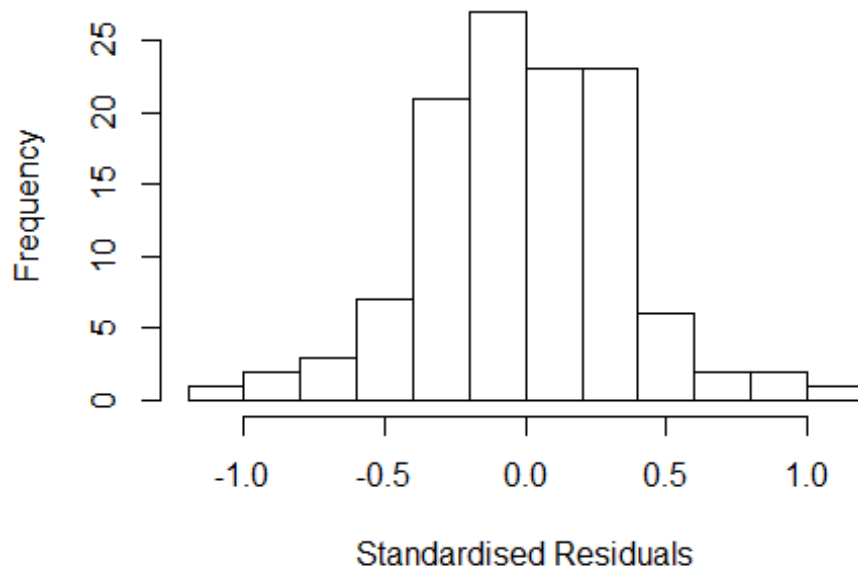
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.96909 -0.20375 -0.00916 0.23559 1.09861
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0543 0.0671 15.72 < 2e-16 ***
## FirstAuthorFemale1 0.1655 0.0728 2.27 0.02522 *
## LastAuthorFemale1 0.1666 0.0820 2.03 0.04503 *
## Year1997 0.3533 0.1540 2.29 0.02386 *
## Year1998 -0.0879 0.1451 -0.61 0.54608
## Year1999 0.0860 0.2263 0.38 0.70482
## Year2000 0.0742 0.0800 0.93 0.35610
## Year2001 0.2547 0.0671 3.80 0.00025 ***
## Year2002 -0.1903 0.0671 -2.84 0.00554 **
## Year2003 -0.0721 0.0886 -0.81 0.41750
## Year2004 0.3247 0.0998 3.25 0.00156 **
## Year2005 0.1066 0.1184 0.90 0.37028
```

```

## Year2006          0.0417      0.1640      0.25  0.80000
## Year2007          0.0261      0.0998      0.26  0.79420
## Year2008         -0.1080      0.1264     -0.85  0.39483
## Year2009          0.0222      0.1190      0.19  0.85259
## Year2010         -0.0926      0.0965     -0.96  0.33963
## Year2011         -0.1271      0.1472     -0.86  0.39011
## Year2012         -0.0852      0.1267     -0.67  0.50304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.153, Adjusted R-squared:  -0.0013
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 108 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.336  0.903   0.961   0.916   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      8.47e-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 4.09 1          2.022
## Year              4.09 16          1.045

```

Residuals from first author



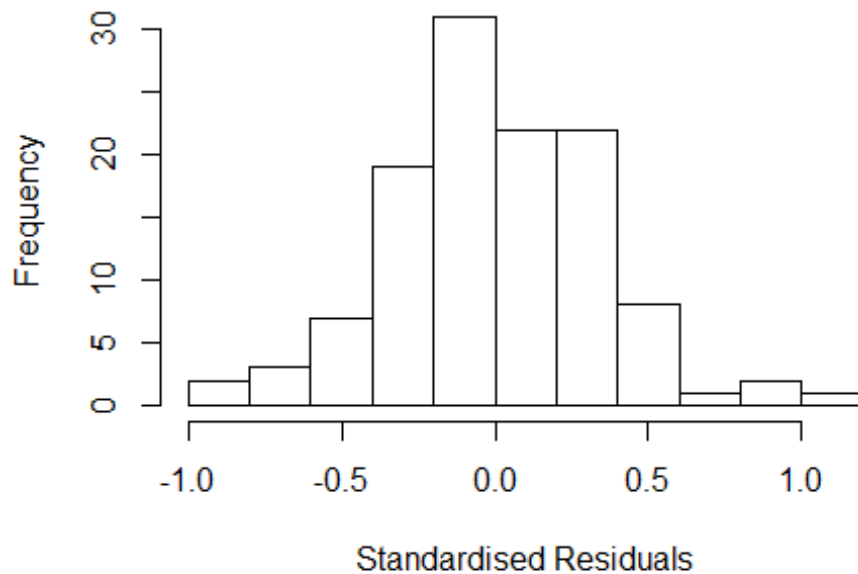
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.03e+00 -2.17e-01 -1.25e-15 2.35e-01 1.06e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09638 0.06791 16.14 <2e-16 ***
## FirstAuthorFemale1 0.16751 0.07523 2.23 0.0282 *
## Year1997 0.37255 0.15659 2.38 0.0193 *
## Year1998 -0.13155 0.14962 -0.88 0.3814
## Year1999 0.04286 0.22735 0.19 0.8508
## Year2000 0.19862 0.06791 2.92 0.0043 **
## Year2001 0.21262 0.06791 3.13 0.0023 **
## Year2002 -0.23238 0.06791 -3.42 0.0009 ***
## Year2003 -0.05844 0.07922 -0.74 0.4624
## Year2004 0.28262 0.10041 2.81 0.0059 **
## Year2005 0.06317 0.12204 0.52 0.6059
## Year2006 -0.00228 0.16636 -0.01 0.9891
```

```

## Year2007          0.02099    0.10387    0.20    0.8403
## Year2008          -0.15197    0.12924   -1.18    0.2424
## Year2009           0.00263    0.13509    0.02    0.9845
## Year2010          -0.08908    0.09902   -0.90    0.3705
## Year2011          -0.10781    0.14993   -0.72    0.4738
## Year2012          -0.06980    0.13408   -0.52    0.6038
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.122, Adjusted R-squared:  -0.0276
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.355  0.905   0.959   0.908   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      8.47e-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 5.698e+13 1      7.548e+06
## Year              5.698e+13 16      2.691e+00

```


Residuals from last author



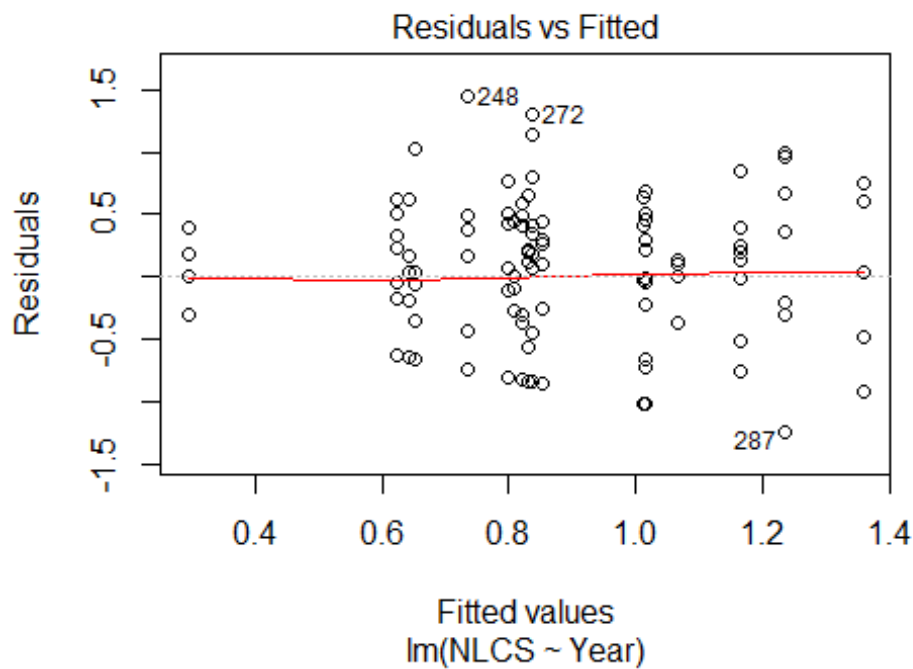
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0528 -0.2022 -0.0124 0.2328 1.0132
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09576 0.06890 15.90 <2e-16 ***
## LastAuthorFemale1 0.16998 0.08772 1.94 0.0555 .
## Year1997 0.37225 0.15693 2.37 0.0196 *
## Year1998 0.00607 0.12232 0.05 0.9605
## Year1999 0.12724 0.32026 0.40 0.6920
## Year2000 0.02926 0.09056 0.32 0.7473
## Year2001 0.21324 0.06890 3.09 0.0026 **
## Year2002 -0.23176 0.06890 -3.36 0.0011 **
## Year2003 -0.11473 0.09215 -1.25 0.2160
## Year2004 0.28324 0.10108 2.80 0.0061 **
## Year2005 0.16959 0.10819 1.57 0.1202
## Year2006 0.11089 0.16576 0.67 0.5050
```

```

## Year2007      0.07001    0.09965    0.70    0.4840
## Year2008     -0.03244    0.12488   -0.26    0.7956
## Year2009      0.04119    0.13774    0.30    0.7655
## Year2010     -0.04684    0.10456   -0.45    0.6551
## Year2011     -0.07882    0.14436   -0.55    0.5863
## Year2012     -0.04293    0.14188   -0.30    0.7628
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.106, Adjusted R-squared:  -0.0456
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 109 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.363  0.891  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      8.47e-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: 118"
## [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
##   19   13   14   21   21   14   15   19   14   22   14   20   23   31   31
## 2011 2012
##   32   28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    6    5    6    6    5    5   10    4   12    7    9   13    8   10
## 2011 2012

```

```
##      9      5
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      5      6      4      6      6      5      5      9      3      11      7      9      12      8      9
## 2011 2012
##      9      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 = 20, df = 16, p-value = 0.2
```



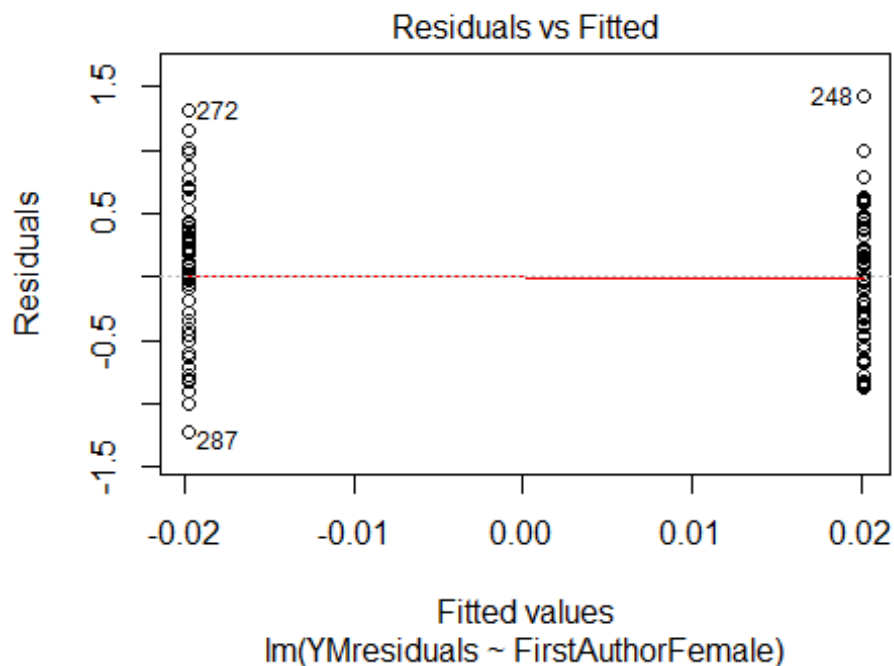
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 4.35587717469286"
## [1] "Male first author team size 2018 geometric mean: 2.05976714390712"

## 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 = 15, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.47722557505166"
## [1] "Male last author team size 2018 geometric mean: 2.44948974278318"

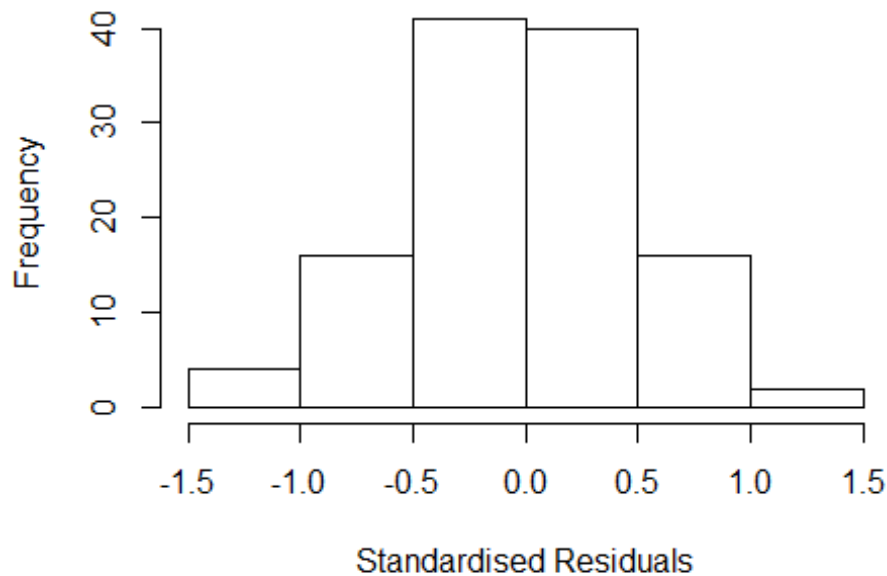
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12, p-value = 0.06
## 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	2.378	1	1.542
LastAuthorFemale	2.561	1	1.600
UniqueAuthors	20.082	4	1.455
Year	78.122	16	1.146

Residuals from first and last author and team size



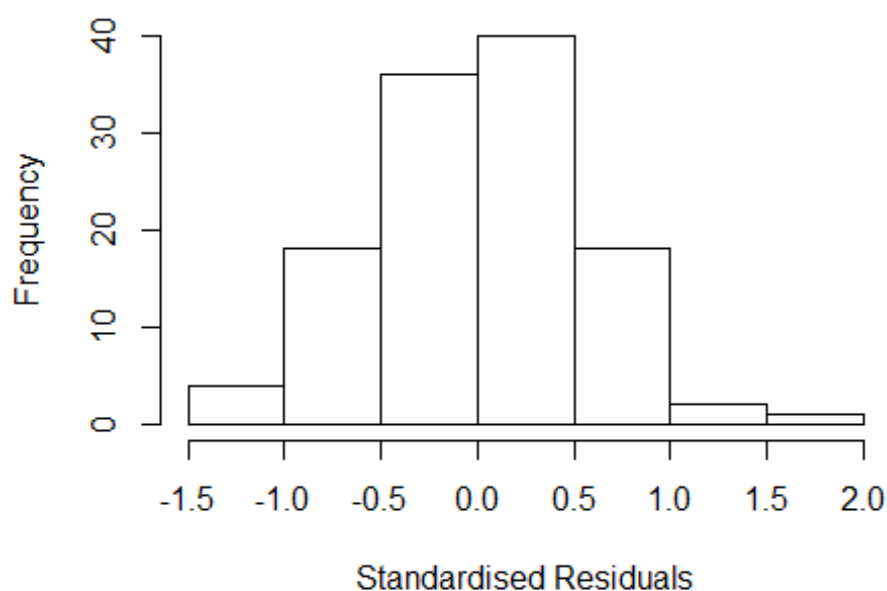
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21696 -0.32787 -0.00555 0.32221 1.21695
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.38368 0.16527 2.32 0.0224 *
## FirstAuthorFemale1 -0.12488 0.12064 -1.04 0.3032
## LastAuthorFemale1 0.09949 0.12706 0.78 0.4356
## UniqueAuthors2 0.67054 0.22676 2.96 0.0039 **
## UniqueAuthors3 0.51549 0.15966 3.23 0.0017 **
## UniqueAuthors4 0.41038 0.19341 2.12 0.0364 *
## UniqueAuthors5 0.84157 0.18887 4.46 2.3e-05 ***
## Year1997 -0.25273 0.17490 -1.44 0.1517
## Year1998 -0.04402 0.20963 -0.21 0.8341
## Year1999 0.04140 0.23588 0.18 0.8610
```

```

## Year2000      0.07319    0.31701    0.23    0.8179
## Year2001      0.13972    0.21923    0.64    0.5254
## Year2002      0.09366    0.26204    0.36    0.7216
## Year2003     -0.38371    0.28120   -1.36    0.1756
## Year2004     -0.02818    0.37682   -0.07    0.9406
## Year2005      0.21982    0.21827    1.01    0.3164
## Year2006     -0.01478    0.21454   -0.07    0.9452
## Year2007     -0.07813    0.36035   -0.22    0.8288
## Year2008     -0.00829    0.34492   -0.02    0.9809
## Year2009      0.51205    0.34257    1.49    0.1383
## Year2010      0.19400    0.28173    0.69    0.4927
## Year2011     -0.11729    0.26327   -0.45    0.6570
## Year2012      0.26393    0.31851    0.83    0.4094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.369, Adjusted R-squared:  0.224
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 106 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.538  0.876  0.951  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      8.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "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  1.897 1      1.377
## Year              3.286 16      1.038

```

Residuals from first and last author



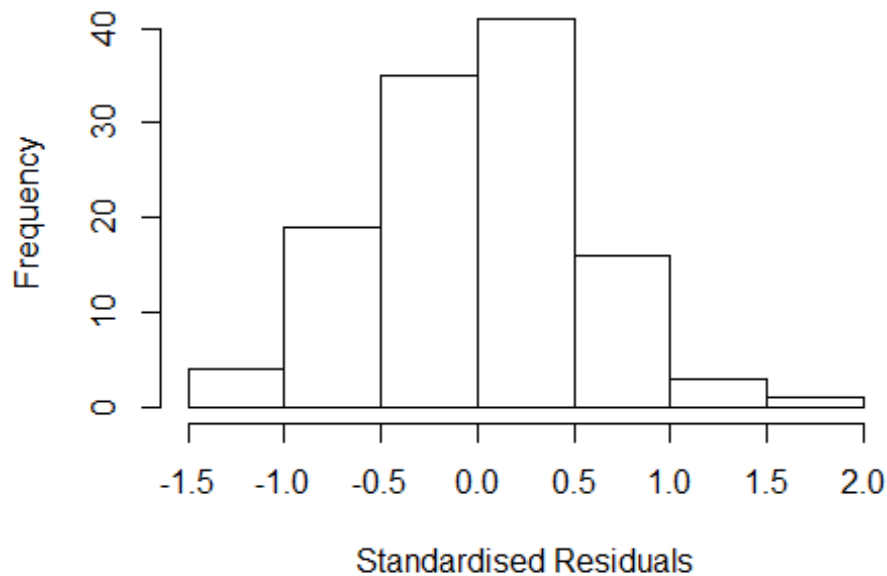
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3931 -0.3700 0.0185 0.3759 1.5642
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.60487 0.25643 2.36 0.020 *
## FirstAuthorFemale1 -0.05011 0.12652 -0.40 0.693
## LastAuthorFemale1 0.09406 0.13614 0.69 0.491
## Year1997 -0.32440 0.27896 -1.16 0.248
## Year1998 0.27864 0.27045 1.03 0.305
## Year1999 0.20837 0.37950 0.55 0.584
## Year2000 0.28029 0.32641 0.86 0.393
## Year2001 0.44869 0.28336 1.58 0.116
## Year2002 0.03776 0.32919 0.11 0.909
## Year2003 -0.06375 0.32205 -0.20 0.843
## Year2004 0.50251 0.61195 0.82 0.414
## Year2005 0.54496 0.29641 1.84 0.069 .
```

```

## Year2006          0.25892    0.31244    0.83    0.409
## Year2007          0.00997    0.34778    0.03    0.977
## Year2008          0.10353    0.40222    0.26    0.797
## Year2009          0.74432    0.54006    1.38    0.171
## Year2010          0.56348    0.31515    1.79    0.077 .
## Year2011          0.20213    0.33158    0.61    0.544
## Year2012          0.73828    0.43795    1.69    0.095 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.205, Adjusted R-squared:  0.0615
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 111 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.374  0.867  0.949  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      8.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.739 1      1.319
## Year              1.739 16      1.017

```


Residuals from first author



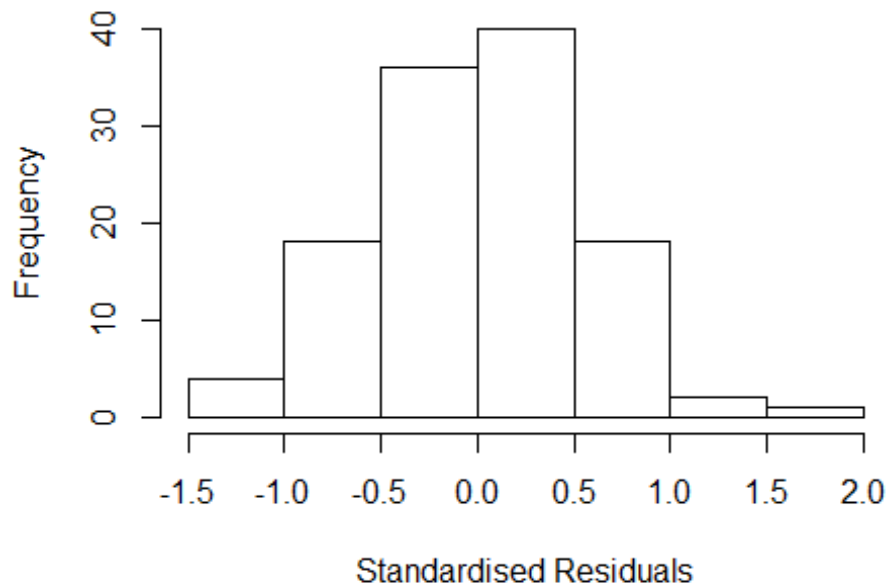
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3250 -0.3719 0.0581 0.3904 1.5065
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6137 0.2689 2.28 0.025 *
## FirstAuthorFemale1 -0.0434 0.1238 -0.35 0.727
## Year1997 -0.3055 0.2813 -1.09 0.280
## Year1998 0.2684 0.2832 0.95 0.346
## Year1999 0.2455 0.3569 0.69 0.493
## Year2000 0.2822 0.3259 0.87 0.389
## Year2001 0.4727 0.2771 1.71 0.091 .
## Year2002 0.0486 0.3300 0.15 0.883
## Year2003 -0.0243 0.3016 -0.08 0.936
## Year2004 0.5255 0.5912 0.89 0.376
## Year2005 0.5266 0.3085 1.71 0.091 .
## Year2006 0.2562 0.3245 0.79 0.432
```

```

## Year2007          0.0587      0.3410      0.17      0.864
## Year2008          0.1522      0.3921      0.39      0.699
## Year2009          0.7546      0.5124      1.47      0.144
## Year2010          0.5740      0.3159      1.82      0.072 .
## Year2011          0.2360      0.3275      0.72      0.473
## Year2012          0.7933      0.4315      1.84      0.069 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.191, Adjusted R-squared:  0.0549
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 110 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.460  0.884  0.955  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      8.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.861 1      1.364
## Year              1.861 16      1.020

```

Residuals from last author



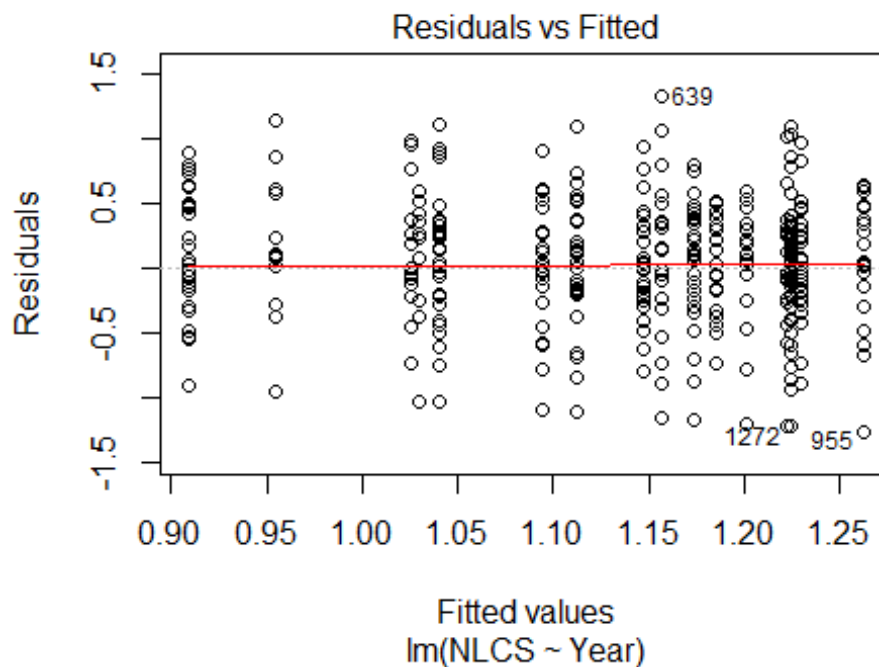
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3941 -0.3679 0.0331 0.4009 1.5909
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.59402 0.24669 2.41 0.018 *
## LastAuthorFemale1 0.08843 0.13543 0.65 0.515
## Year1997 -0.32810 0.27692 -1.18 0.239
## Year1998 0.27675 0.26919 1.03 0.306
## Year1999 0.19345 0.37824 0.51 0.610
## Year2000 0.27197 0.32057 0.85 0.398
## Year2001 0.44067 0.28128 1.57 0.120
## Year2002 0.03065 0.32319 0.09 0.925
## Year2003 -0.05460 0.31637 -0.17 0.863
## Year2004 0.50817 0.61973 0.82 0.414
## Year2005 0.51936 0.29921 1.74 0.086 .
## Year2006 0.24891 0.31278 0.80 0.428
```

```

## Year2007          -0.00588      0.34364    -0.02      0.986
## Year2008           0.10642      0.40005      0.27      0.791
## Year2009           0.71169      0.53135      1.34      0.183
## Year2010           0.55469      0.31237      1.78      0.079 .
## Year2011           0.17427      0.32590      0.53      0.594
## Year2012           0.71705      0.43468      1.65      0.102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.546
## Multiple R-squared:  0.201, Adjusted R-squared:  0.0669
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 109 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.376  0.865   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      8.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 119"
## [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
##   71   67   73   41   62   58   80   64   53   64   67   99   74   74   78
## 2011 2012
##   66   92
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32   30   26   18   18   16   12   19   25   20   32   28   20   26   31
## 2011 2012

```

```
## 27 35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 28 20 17 17 14 10 15 23 14 19 21 16 19 27
## 2011 2012
## 20 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 = 15, df = 16, p-value = 0.5
```



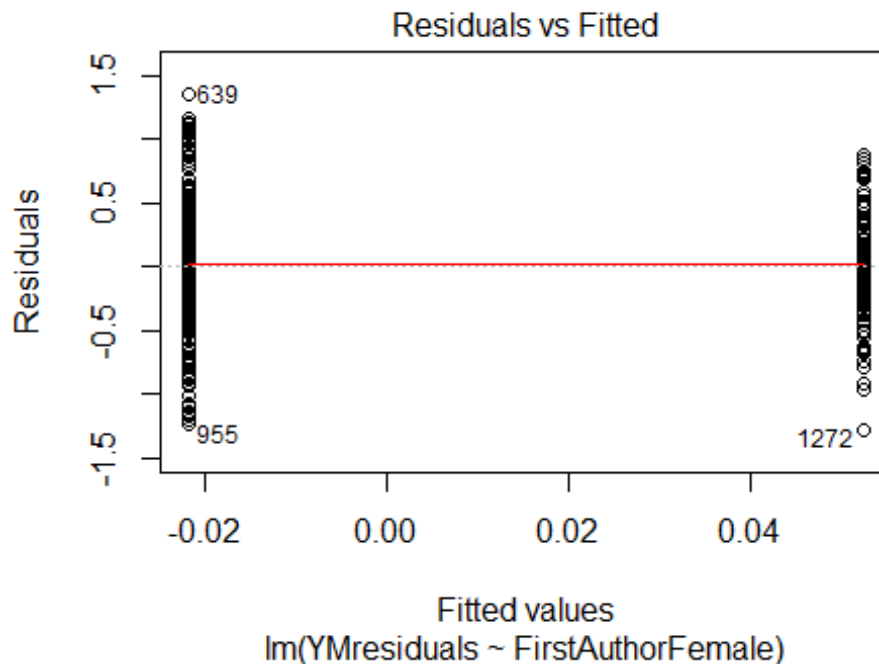
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.2, df = 1, p-value = 0.01

## [1] "Female first author team size 2018 geometric mean: 3.69702406876445"
## [1] "Male first author team size 2018 geometric mean: 3.88742767200806"

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

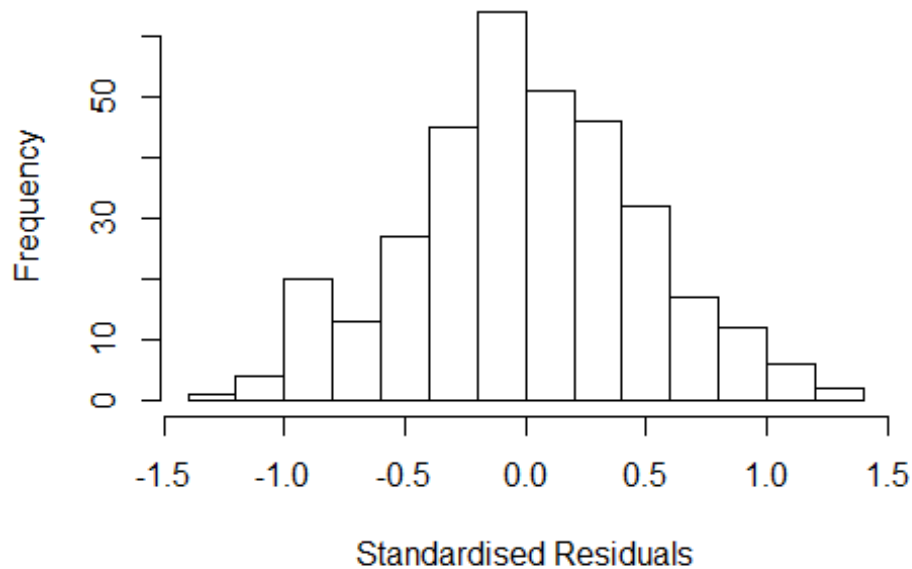
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 66, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.21336383940064"
## [1] "Male last author team size 2018 geometric mean: 4.27393573527291"

## 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 = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.751  1      1.323
## LastAuthorFemale  1.925  1      1.387
## UniqueAuthors    2.901  4      1.142
## Year              3.574 16      1.041
```

Residuals from first and last author and team size



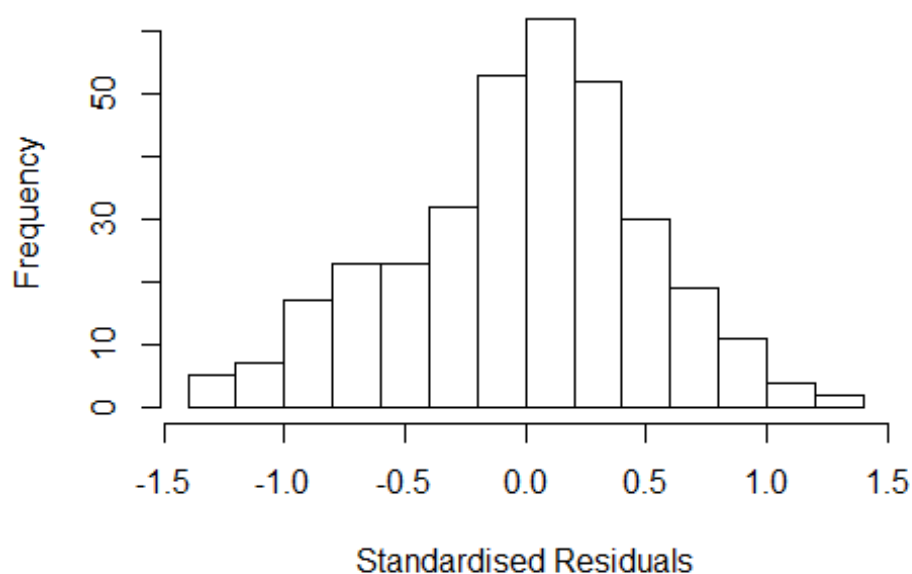
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29143 -0.31015 -0.00934 0.31667 1.27544
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0817 0.0825 13.11 < 2e-16 ***
## FirstAuthorFemale1 -0.0462 0.0710 -0.65 0.51574
## LastAuthorFemale1 0.0193 0.0933 0.21 0.83663
## UniqueAuthors2 0.1600 0.0886 1.81 0.07200 .
## UniqueAuthors3 0.3612 0.0827 4.37 1.7e-05 ***
## UniqueAuthors4 0.4513 0.0931 4.85 2.0e-06 ***
## UniqueAuthors5 0.3791 0.0982 3.86 0.00014 ***
## Year1997 -0.0473 0.1336 -0.35 0.72326
## Year1998 0.0929 0.1273 0.73 0.46625
## Year1999 -0.1998 0.1583 -1.26 0.20767
```

```

## Year2000          -0.0678      0.1090    -0.62  0.53413
## Year2001          -0.4414      0.1682    -2.62  0.00912 **
## Year2002          -0.1167      0.2014    -0.58  0.56276
## Year2003          -0.1459      0.1301    -1.12  0.26289
## Year2004          -0.1732      0.1468    -1.18  0.23895
## Year2005           0.0497      0.1408      0.35  0.72422
## Year2006          -0.2483      0.1414    -1.76  0.07992 .
## Year2007          -0.2009      0.1873    -1.07  0.28426
## Year2008          -0.1078      0.1472    -0.73  0.46435
## Year2009          -0.2728      0.1253    -2.18  0.03025 *
## Year2010          -0.2575      0.1211    -2.13  0.03429 *
## Year2011          -0.0789      0.1338    -0.59  0.55584
## Year2012          -0.3886      0.1337    -2.91  0.00391 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.16,   Adjusted R-squared:  0.102
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 303 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.422  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          2.94e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "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  1.854 1          1.361
## Year              1.598 16          1.015

```


Residuals from first and last author



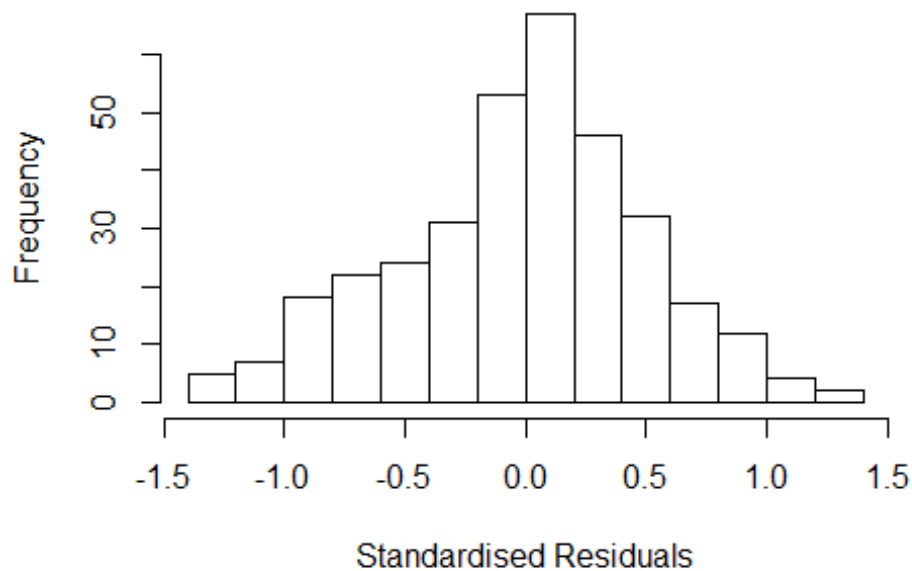
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3366 -0.3042 0.0206 0.3156 1.3382
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.219089 0.072180 16.89 <2e-16 ***
## FirstAuthorFemale1 0.084738 0.069689 1.22 0.2249
## LastAuthorFemale1 -0.063403 0.098850 -0.64 0.5217
## Year1997 -0.012688 0.128296 -0.10 0.9213
## Year1998 0.050870 0.127849 0.40 0.6910
## Year1999 -0.186737 0.155653 -1.20 0.2311
## Year2000 -0.026432 0.120515 -0.22 0.8265
## Year2001 -0.398845 0.178748 -2.23 0.0263 *
## Year2002 -0.152469 0.207062 -0.74 0.4621
## Year2003 -0.064596 0.143061 -0.45 0.6519
## Year2004 -0.054624 0.162398 -0.34 0.7368
## Year2005 0.117520 0.156707 0.75 0.4538
```

```

## Year2006      -0.171287    0.160926   -1.06    0.2880
## Year2007      -0.106627    0.158197   -0.67    0.5008
## Year2008      -0.000886    0.162161   -0.01    0.9956
## Year2009      -0.227120    0.124201   -1.83    0.0684 .
## Year2010      -0.100854    0.110297   -0.91    0.3612
## Year2011      -0.038883    0.137686   -0.28    0.7778
## Year2012      -0.359965    0.136240   -2.64    0.0086 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.0626, Adjusted R-squared:  0.01
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 308 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.452  0.845   0.951   0.901   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      2.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.124 1      1.060
## Year              1.124 16      1.004

```

Residuals from first author



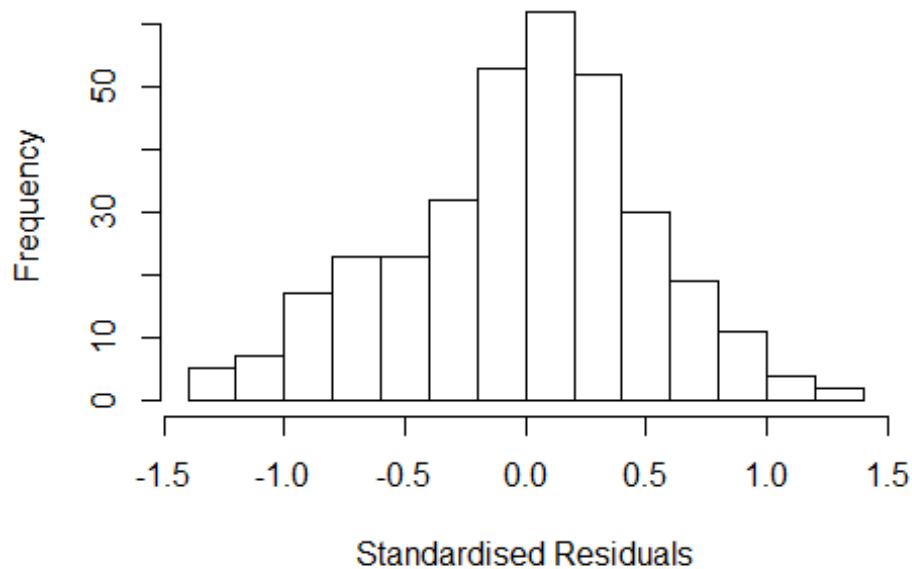
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3286 -0.3168 0.0273 0.3198 1.3193
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21377 0.07331 16.56 <2e-16 ***
## FirstAuthorFemale1 0.05831 0.06035 0.97 0.335
## Year1997 -0.00321 0.12797 -0.03 0.980
## Year1998 0.05389 0.12760 0.42 0.673
## Year1999 -0.17324 0.15458 -1.12 0.263
## Year2000 -0.02651 0.11977 -0.22 0.825
## Year2001 -0.38584 0.17510 -2.20 0.028 *
## Year2002 -0.13633 0.20480 -0.67 0.506
## Year2003 -0.07062 0.14172 -0.50 0.619
## Year2004 -0.05006 0.16365 -0.31 0.760
## Year2005 0.11480 0.15894 0.72 0.471
## Year2006 -0.17282 0.16183 -1.07 0.286
```

```

## Year2007      -0.09467    0.15955   -0.59    0.553
## Year2008      0.00326    0.16066    0.02    0.984
## Year2009     -0.22390    0.12388   -1.81    0.072 .
## Year2010     -0.10375    0.10908   -0.95    0.342
## Year2011     -0.03913    0.13851   -0.28    0.778
## Year2012     -0.36884    0.13580   -2.72    0.007 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0614, Adjusted R-squared:  0.0118
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 310 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.455  0.848  0.951  0.900  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      2.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.375 1      1.173
## Year              1.375 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
## -1.344 -0.288 0.019 0.328 1.304
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22833 0.07204 17.05 <2e-16 ***
## LastAuthorFemale1 -0.00776 0.08502 -0.09 0.927
## Year1997 -0.00582 0.13029 -0.04 0.964
## Year1998 0.05979 0.12601 0.47 0.635
## Year1999 -0.16943 0.15515 -1.09 0.276
## Year2000 -0.01400 0.11918 -0.12 0.907
## Year2001 -0.38906 0.17772 -2.19 0.029 *
## Year2002 -0.12854 0.20252 -0.63 0.526
## Year2003 -0.05533 0.14400 -0.38 0.701
## Year2004 -0.04886 0.16229 -0.30 0.764
## Year2005 0.11533 0.15548 0.74 0.459
## Year2006 -0.17282 0.16357 -1.06 0.292
```

```

## Year2007          -0.09547      0.16166    -0.59      0.555
## Year2008          -0.00577      0.16087    -0.04      0.971
## Year2009          -0.21425      0.12597    -1.70      0.090 .
## Year2010          -0.08977      0.11139    -0.81      0.421
## Year2011          -0.03615      0.13677    -0.26      0.792
## Year2012          -0.35202      0.14045    -2.51      0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0591, Adjusted R-squared:  0.00939
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 311 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.444  0.845   0.953   0.900   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      2.94e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 340"
## [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 2006 2007 2008 2009 2010 2011 2012
##    2    3    2    1    1    2    2    3    4    6
##
## 1996 1997 1998 2006 2007 2008 2009 2010 2011 2012
##    0    0    0    0    0    0    0    2    3    1
##
## 1996 1997 1998 2006 2007 2008 2009 2010 2011 2012
##    0    0    0    0    0    0    0    2    2    1

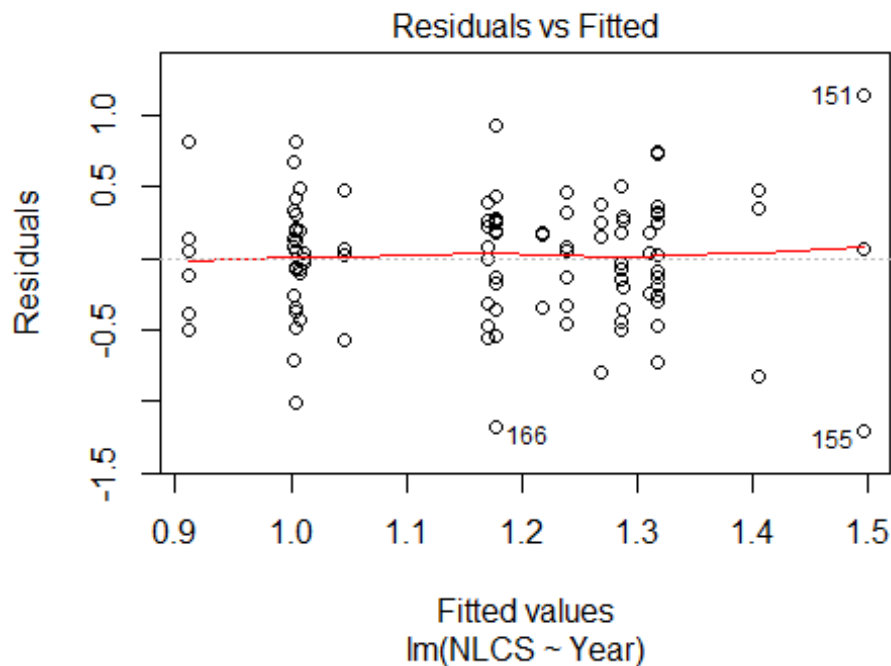
```

```

## [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"
## [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: 5"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3503"
## [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: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [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 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
## 10 5 14 9 10 5 12 9 11 5 11 9 6 16 19
## 2011 2012
## 18 25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 3 4 6 4 3 7 8 7 3 4 3 4 13 14
## 2011 2012
## 9 18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 6 1 3 5 4 3 7 5 7 3 4 2 3 10 13

```

```
## 2011 2012
## 9 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 = 23, df = 16, p-value = 0.1
```



```
##
## 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: 6"
## [1] "Male first author team size 2018 geometric mean: 3.97920782685131"

## 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 = 0.2
```

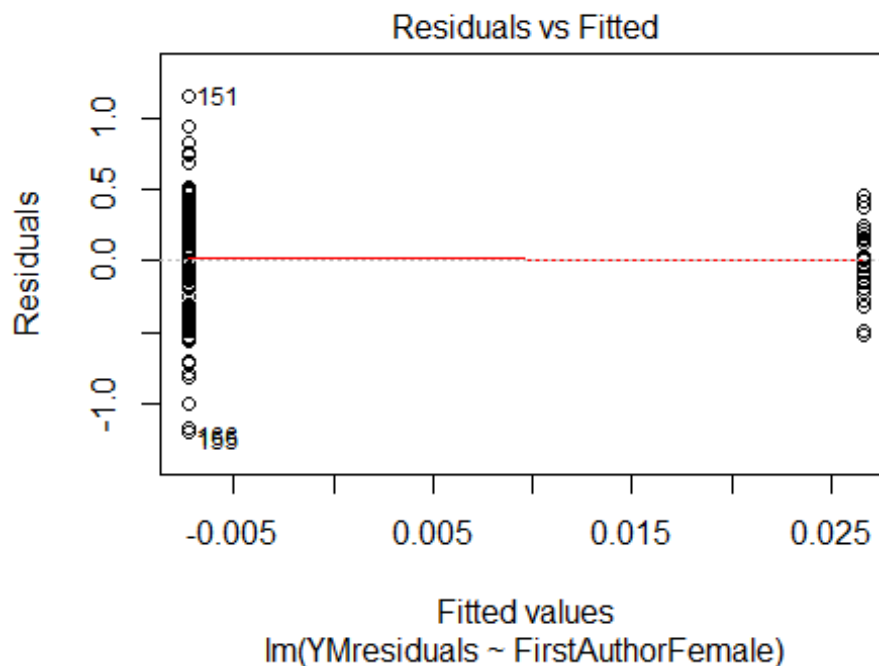


```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4"
## [1] "Male last author team size 2018 geometric mean: 4.40754408739457"

## 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, 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 outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2007'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.339e+14  1      1.157e+07
## LastAuthorFemale  3.937e+00  1      1.984e+00
## UniqueAuthors    2.446e+16  4      1.118e+02
## Year              1.210e+17 16      3.419e+00
```

```

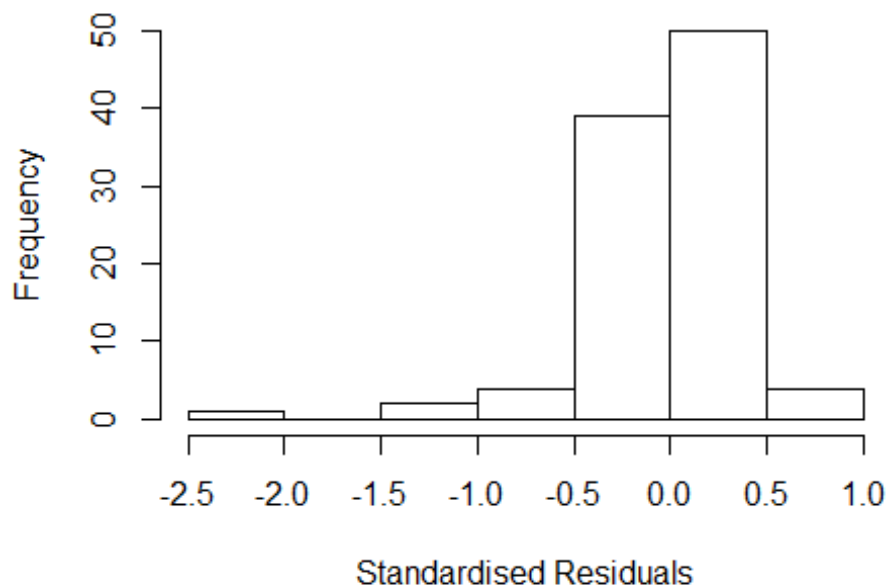
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.4099 -0.2748 0.0134 0.2047 0.9510
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8246 0.2310 3.57 0.00062 ***
## FirstAuthorFemale1 -0.0327 0.0959 -0.34 0.73361
## LastAuthorFemale1 0.1541 0.1623 0.95 0.34554
## UniqueAuthors2 0.1188 0.2185 0.54 0.58824
## UniqueAuthors3 0.1028 0.1638 0.63 0.53208
## UniqueAuthors4 0.0399 0.1812 0.22 0.82636
## UniqueAuthors5 -0.0363 0.1665 -0.22 0.82785
## Year1997 0.4963 0.2478 2.00 0.04868 *
## Year1998 0.5471 0.2612 2.09 0.03952 *
## Year1999 0.1172 0.2522 0.46 0.64356
## Year2000 0.4183 0.2518 1.66 0.10078
## Year2001 0.1465 0.2239 0.65 0.51483
## Year2002 0.4125 0.2547 1.62 0.10944
## Year2003 0.4786 0.3049 1.57 0.12062
## Year2004 0.1514 0.2838 0.53 0.59529
## Year2005 0.8592 0.4306 2.00 0.04955 *
## Year2006 0.1099 0.3240 0.34 0.73541
## Year2007 1.7735 0.2202 8.06 8e-12 ***
## Year2008 0.5078 0.5081 1.00 0.32066
## Year2009 0.2915 0.3682 0.79 0.43093
## Year2010 0.1858 0.2301 0.81 0.42205
## Year2011 0.2386 0.2846 0.84 0.40439
## Year2012 0.3958 0.2353 1.68 0.09655 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared: 0.32, Adjusted R-squared: 0.125
## Convergence in 34 IRWLS iterations
##
## Robustness weights:
## observation 50 is an outlier with |weight| = 0 ( < 0.001);
## 11 weights are ~= 1. The remaining 88 ones are summarized as

```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    0.180  0.875  0.945  0.893  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          1.00e-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"

## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2007'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```

Residuals from first and last author and team size



```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 5.502e+13  1      7.418e+06
```

```

## LastAuthorFemale  2.719e+00  1      1.649e+00
## Year              1.487e+14 16      2.773e+00

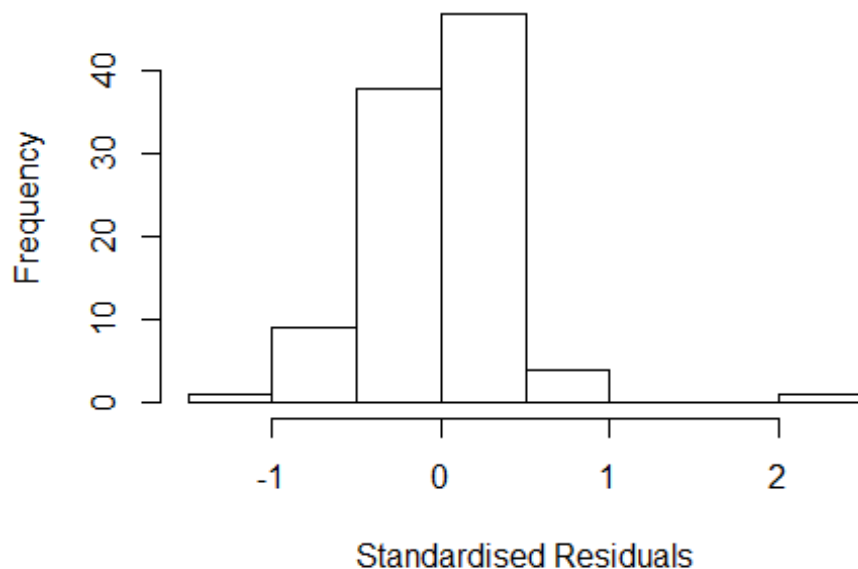
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##   Min       1Q   Median       3Q      Max
## -1.209 -0.224  0.042  0.227  2.347
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8710     0.2174   4.01 0.00014 ***
## FirstAuthorFemale1 -0.0436     0.0797  -0.55 0.58610
## LastAuthorFemale1  0.1722     0.1527   1.13 0.26271
## Year1997         0.5635     0.1941   2.90 0.00475 **
## Year1998         0.5421     0.2601   2.08 0.04029 *
## Year1999         0.1654     0.2524   0.66 0.51425
## Year2000         0.4242     0.2375   1.79 0.07783 .
## Year2001         0.1406     0.2181   0.64 0.52088
## Year2002         0.3749     0.2453   1.53 0.13034
## Year2003         0.4743     0.2952   1.61 0.11200
## Year2004         0.1445     0.2747   0.53 0.60034
## Year2005         0.6760     0.5023   1.35 0.18211
## Year2006         0.1600     0.2847   0.56 0.57569
## Year2007        -0.5800     0.2174  -2.67 0.00920 **
## Year2008         0.4350     0.4524   0.96 0.33920
## Year2009         0.3382     0.2961   1.14 0.25680
## Year2010         0.2169     0.2276   0.95 0.34343
## Year2011         0.2578     0.2650   0.97 0.33363
## Year2012         0.3881     0.2218   1.75 0.08393 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.396
## Multiple R-squared:  0.195, Adjusted R-squared:  0.0158
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 49 is an outlier with |weight| = 0 ( < 0.001);
## 9 weights are ~1. The remaining 90 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.330  0.895  0.955  0.919  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"

## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2007'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```

Residuals from first and last author



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.399e+14 1 1.183e+07
## Year 1.399e+14 16 2.767e+00

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
```

```

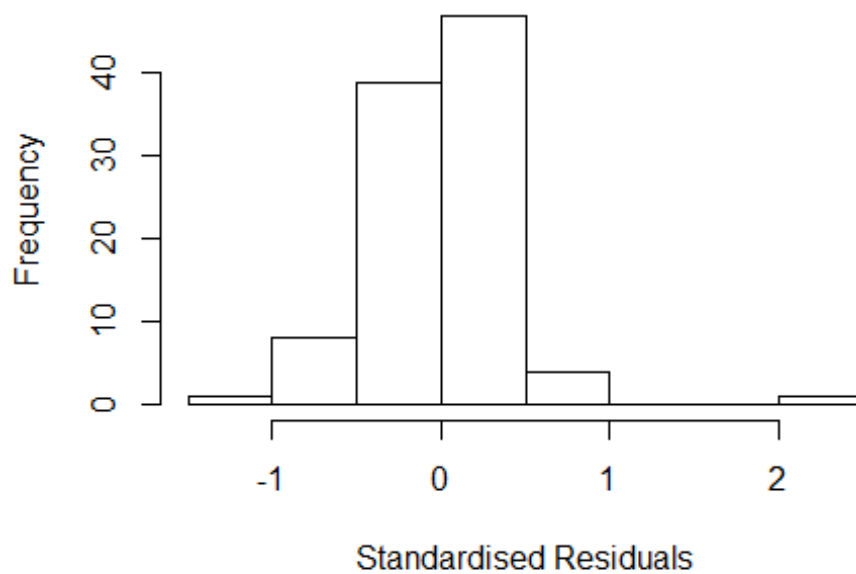
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2644 -0.2486  0.0359  0.2087  2.3470
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.86462    0.19820   4.36 3.7e-05 ***
## FirstAuthorFemale1 0.00236    0.09241   0.03  0.9797
## Year1997        0.52402    0.18251   2.87  0.0052 **
## Year1998        0.54875    0.24519   2.24  0.0279 *
## Year1999        0.16087    0.23849   0.67  0.5019
## Year2000        0.44660    0.20174   2.21  0.0296 *
## Year2001        0.14704    0.19904   0.74  0.4622
## Year2002        0.37359    0.22923   1.63  0.1070
## Year2003        0.48074    0.28206   1.70  0.0921 .
## Year2004        0.14234    0.26068   0.55  0.5865
## Year2005        0.69373    0.50302   1.38  0.1716
## Year2006        0.19528    0.28482   0.69  0.4949
## Year2007       -0.57362    0.19820  -2.89  0.0049 **
## Year2008        0.44788    0.45140   0.99  0.3240
## Year2009        0.39980    0.27027   1.48  0.1429
## Year2010        0.21046    0.21052   1.00  0.3204
## Year2011        0.31813    0.23228   1.37  0.1745
## Year2012        0.39703    0.20198   1.97  0.0527 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.192, Adjusted R-squared:  0.024
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 49 is an outlier with |weight| = 0 ( < 0.001);
## 11 weights are ~= 1. The remaining 88 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.263  0.885  0.958  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      1.00e-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"

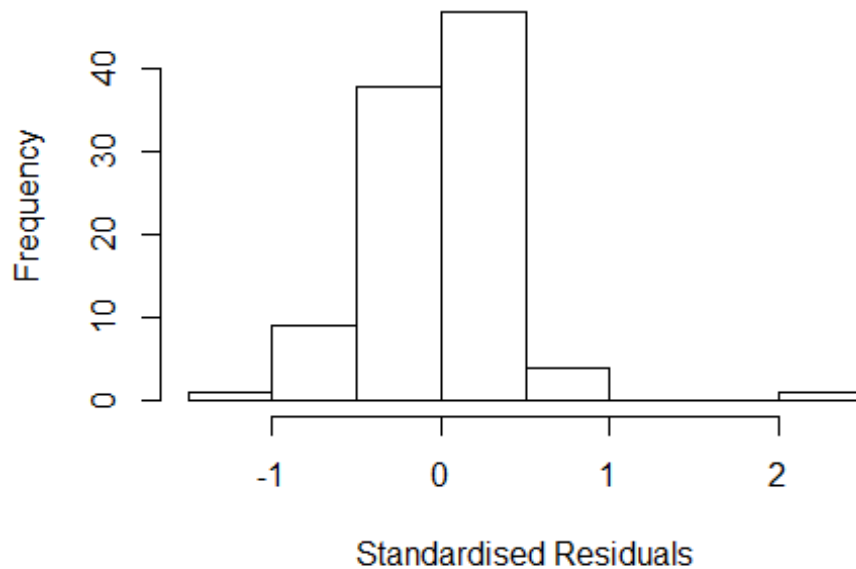
## Warning in outlierStats(ret, x, control): Detected possible local
breakdown of SM-estimate in coefficient 'Year2007'.
## Use lmrob argument 'setting="KS2014"' to avoid this problem.
```

Residuals from first author



```
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 2.622 1 1.619
## Year 2.622 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
## -1.1950 -0.2439 0.0467 0.2174 2.3470
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.848 0.196 4.33 4.2e-05 ***
## LastAuthorFemale1 0.147 0.153 0.96 0.3396
## Year1997 0.543 0.196 2.77 0.0069 **
## Year1998 0.565 0.243 2.33 0.0223 *
## Year1999 0.178 0.244 0.73 0.4670
## Year2000 0.432 0.230 1.88 0.0639 .
## Year2001 0.164 0.197 0.83 0.4073
## Year2002 0.391 0.233 1.68 0.0972 .
## Year2003 0.498 0.280 1.78 0.0794 .
## Year2004 0.159 0.265 0.60 0.5484
## Year2005 0.704 0.492 1.43 0.1561
## Year2006 0.179 0.270 0.66 0.5085
```



```

## Year2007          -0.557      0.196   -2.84   0.0056 **
## Year2008           0.461      0.443    1.04   0.3010
## Year2009           0.347      0.292    1.19   0.2377
## Year2010           0.228      0.219    1.04   0.3006
## Year2011           0.280      0.250    1.12   0.2670
## Year2012           0.404      0.208    1.94   0.0552 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.196, Adjusted R-squared:  0.0291
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 49 is an outlier with |weight| = 0 ( < 0.001);
## 8 weights are ~= 1. The remaining 91 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.332  0.893  0.957  0.919  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] "Sample size for the above analysis: 100"
## [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]"
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    4    7    3    3    2    3    5    2    3    4    3    4    8    3
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    3    0    2    2    1    2    1    3    3    1    2    6    3
##
## 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012

```

```

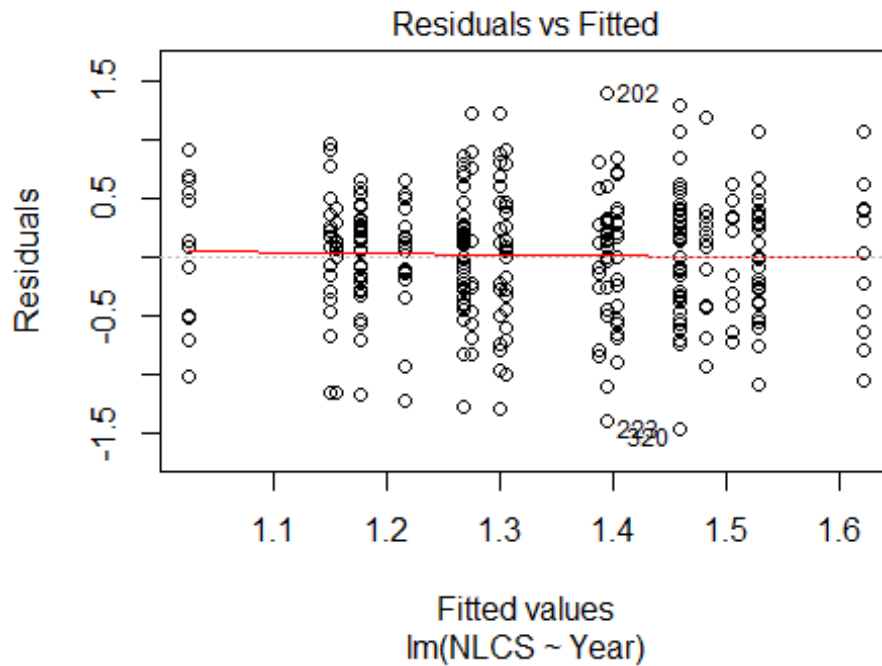
##      1      3      0      1      2      1      1      0      0      2      0      2      4      2
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 5.47722557505166"
## [1] "Male first author team size 2018 geometric mean: 3"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3"
## [1] "Male last author team size 2018 geometric mean: 5.47722557505166"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, 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"
## [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: 19"
## [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 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##      4      2      1     11      7      6      4      1      1      2      1      2      7      6      4
##
## 1996 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##      1      1      0      3      2      4      1      1      0      2      1      0      4      3      1
##
## 1996 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##      1      1      0      3      2      4      1      1      0      2      1      0      4      2      0
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 9"
## [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: 22"
## [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
## 13 19 27 26 20 30 32 31 23 34 48 42 48 14 11
## 2011 2012
## 16 12
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 12 16 17 13 18 17 22 19 29 41 36 38 11 10
## 2011 2012
## 13 10
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 11 15 16 11 18 15 18 18 26 34 32 36 10 10
## 2011 2012
## 12 9
## [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

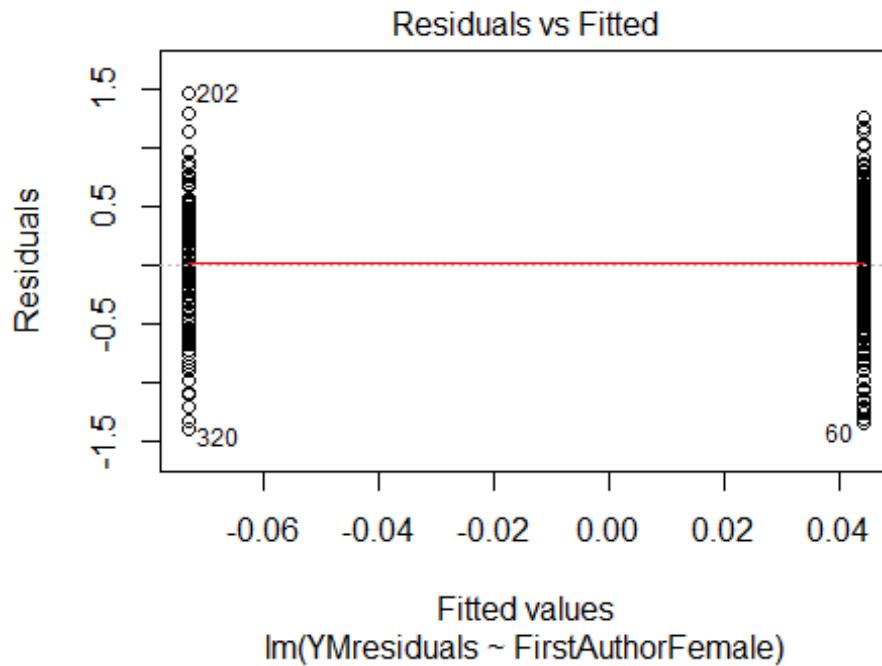
```



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

## [1] "Female first author team size 2018 geometric mean: 1.84218488139299"
## [1] "Male first author team size 2018 geometric mean: 6"

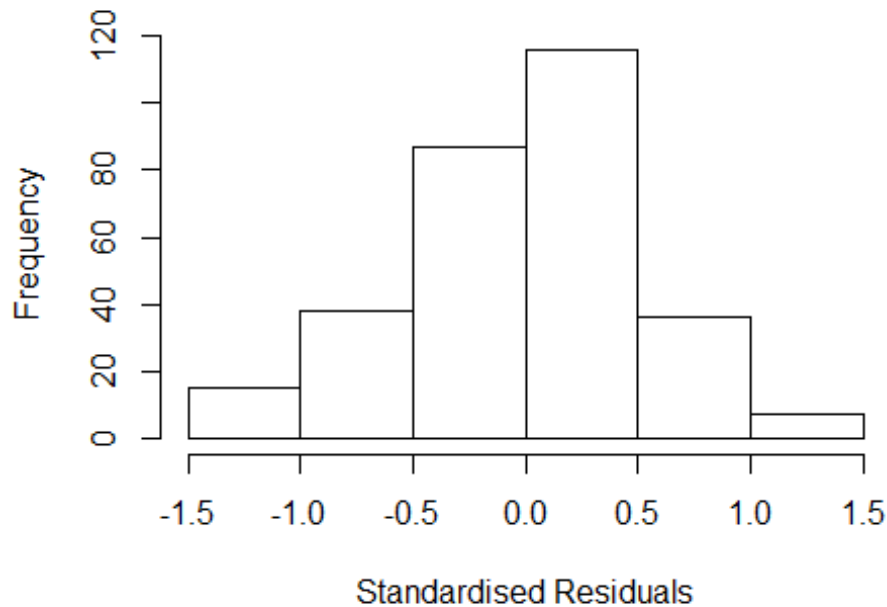
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0.5, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.39492974229681"
## [1] "Male last 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	2.208	1	1.486
LastAuthorFemale	2.041	1	1.429
UniqueAuthors	2.656	4	1.130
Year	5.736	16	1.056

Residuals from first and last author and team size



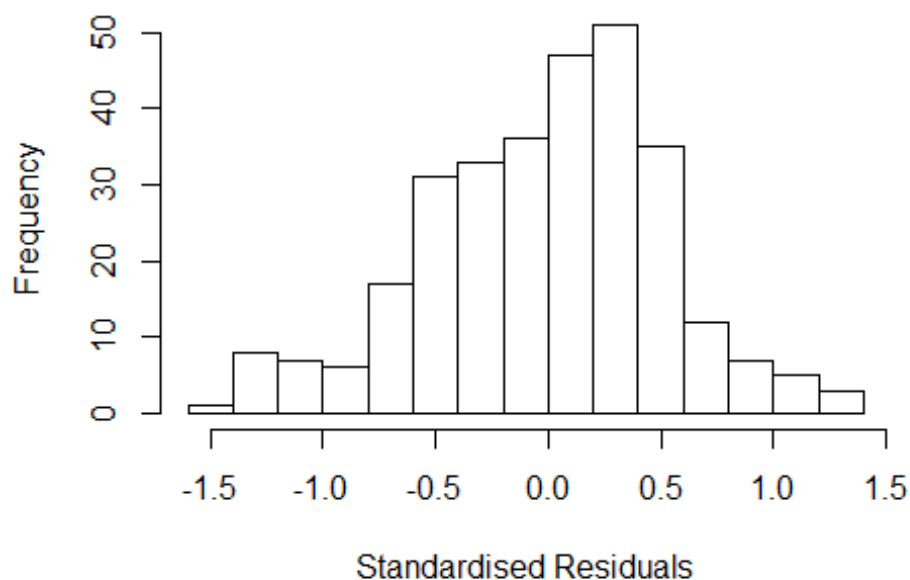
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4269 -0.3524 0.0461 0.3523 1.4692
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12026 0.13248 8.46 1.6e-15 ***
## FirstAuthorFemale1 0.16554 0.08020 2.06 0.0399 *
## LastAuthorFemale1 -0.06789 0.07935 -0.86 0.3929
## UniqueAuthors2 0.18864 0.08550 2.21 0.0282 *
## UniqueAuthors3 0.20532 0.09759 2.10 0.0363 *
## UniqueAuthors4 0.34611 0.10975 3.15 0.0018 **
## UniqueAuthors5 0.33686 0.10639 3.17 0.0017 **
## Year1997 0.41587 0.22524 1.85 0.0659 .
## Year1998 -0.00523 0.25004 -0.02 0.9833
## Year1999 0.04992 0.19736 0.25 0.8005
```

```

## Year2000      -0.40978    0.23636   -1.73    0.0841 .
## Year2001      0.02040    0.15761    0.13    0.8971
## Year2002     -0.11329    0.19756   -0.57    0.5668
## Year2003      0.09678    0.18580    0.52    0.6029
## Year2004      0.02616    0.20518    0.13    0.8987
## Year2005      0.07729    0.15523    0.50    0.6189
## Year2006     -0.00663    0.16220   -0.04    0.9674
## Year2007     -0.20033    0.14822   -1.35    0.1776
## Year2008     -0.10330    0.15620   -0.66    0.5090
## Year2009      0.28151    0.19625    1.43    0.1526
## Year2010      0.10014    0.21379    0.47    0.6399
## Year2011      0.11373    0.22946    0.50    0.6205
## Year2012     -0.08743    0.35512   -0.25    0.8057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.144, Adjusted R-squared:  0.0757
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 268 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.363  0.881  0.943  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      3.34e-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 2.128 1      1.459
## LastAuthorFemale 2.005 1      1.416
## Year      2.390 16      1.028

```

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
## -1.4232 -0.3652 0.0473 0.3222 1.3678
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2384 0.1170 10.59 <2e-16 ***
## FirstAuthorFemale1 0.1803 0.0820 2.20 0.029 *
## LastAuthorFemale1 -0.0854 0.0823 -1.04 0.300
## Year1997 0.4314 0.2324 1.86 0.064 .
## Year1998 -0.0356 0.2450 -0.15 0.885
## Year1999 0.0126 0.1822 0.07 0.945
## Year2000 -0.4414 0.2458 -1.80 0.074 .
## Year2001 -0.0207 0.1534 -0.13 0.893
## Year2002 -0.0732 0.2263 -0.32 0.747
## Year2003 0.1848 0.1802 1.03 0.306
## Year2004 0.0661 0.1977 0.33 0.738
## Year2005 0.1824 0.1510 1.21 0.228
```

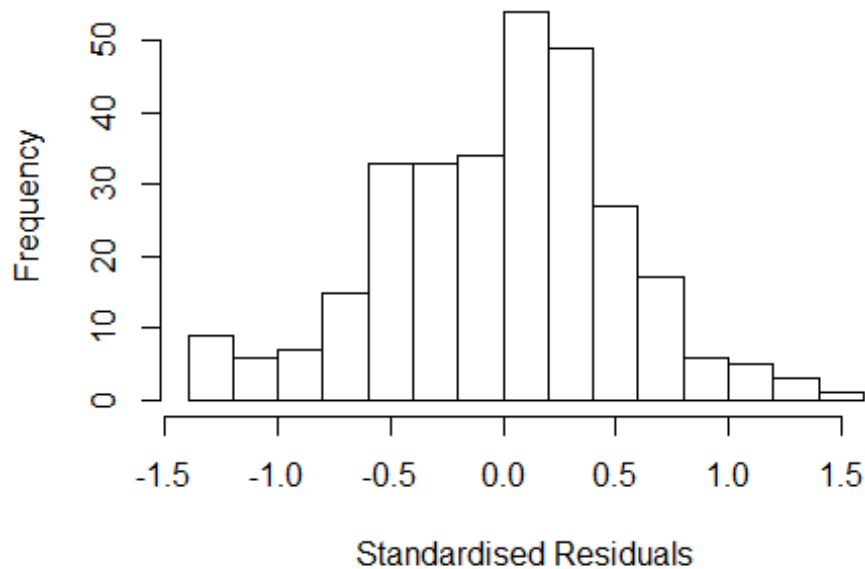


```

## Year2006          0.0983      0.1583      0.62      0.535
## Year2007         -0.1570      0.1457     -1.08      0.282
## Year2008         -0.0371      0.1550     -0.24      0.811
## Year2009          0.2578      0.2022      1.27      0.203
## Year2010          0.1057      0.2224      0.48      0.635
## Year2011          0.1978      0.2227      0.89      0.375
## Year2012         -0.0847      0.3454     -0.25      0.807
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.0991, Adjusted R-squared:  0.0412
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 278 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.418  0.887  0.954  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      3.34e-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 1.651 1      1.285
## Year              1.651 16      1.016

```

Residuals from first author



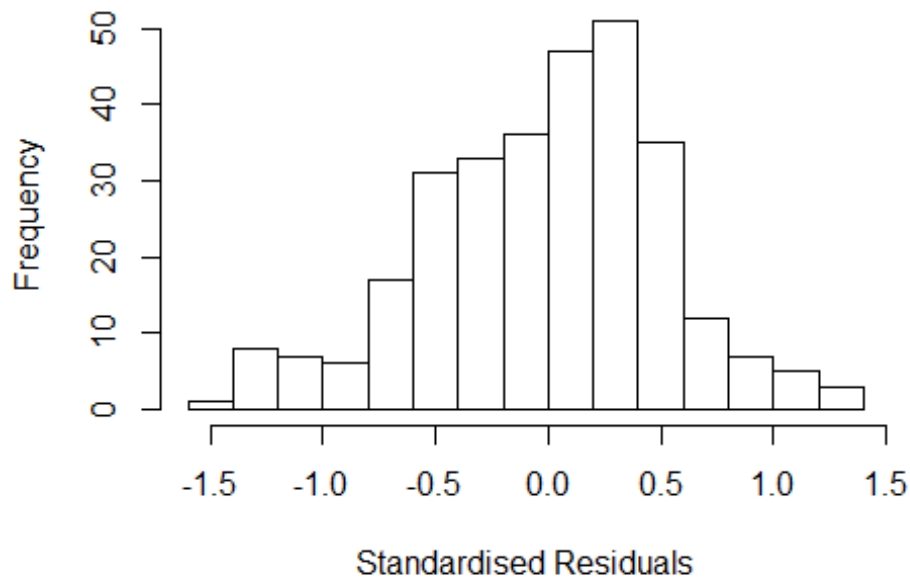
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3836 -0.3545 0.0439 0.3254 1.4074
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2095 0.1118 10.82 <2e-16 ***
## FirstAuthorFemale1 0.1359 0.0720 1.89 0.060 .
## Year1997 0.4288 0.2391 1.79 0.074 .
## Year1998 -0.0304 0.2544 -0.12 0.905
## Year1999 0.0282 0.1790 0.16 0.875
## Year2000 -0.4279 0.2585 -1.66 0.099 .
## Year2001 -0.0239 0.1516 -0.16 0.875
## Year2002 -0.0664 0.2290 -0.29 0.772
## Year2003 0.1741 0.1760 0.99 0.323
## Year2004 0.0873 0.1888 0.46 0.644
## Year2005 0.1989 0.1490 1.33 0.183
## Year2006 0.0982 0.1565 0.63 0.531
```

```

## Year2007          -0.1383      0.1422   -0.97    0.332
## Year2008          -0.0212      0.1516   -0.14    0.889
## Year2009           0.2573      0.2019    1.27    0.204
## Year2010           0.0978      0.2148    0.46    0.649
## Year2011           0.1729      0.2104    0.82    0.412
## Year2012          -0.0712      0.3632   -0.20    0.845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0959, Adjusted R-squared:  0.0412
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 278 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.424  0.879   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      3.34e-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 1.49 1          1.221
## Year              1.49 16          1.013

```

Residuals from last author



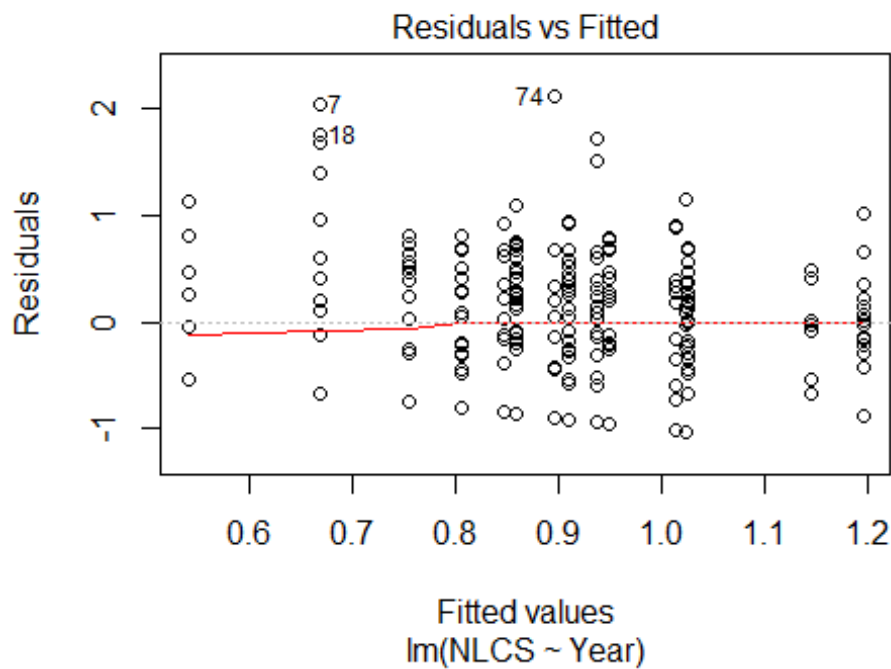
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4675 -0.3508  0.0407  0.3106  1.3648
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21722    0.12172   10.00  <2e-16 ***
## LastAuthorFemale1 0.00673    0.07225    0.09   0.926
## Year1997        0.49052    0.23821    2.06   0.040 *
## Year1998        0.01242    0.26547    0.05   0.963
## Year1999        0.08266    0.18765    0.44   0.660
## Year2000       -0.33677    0.24788   -1.36   0.175
## Year2001        0.03392    0.15818    0.21   0.830
## Year2002       -0.02169    0.22781   -0.10   0.924
## Year2003        0.25033    0.17547    1.43   0.155
## Year2004        0.17453    0.18578    0.94   0.348
## Year2005        0.26875    0.15089    1.78   0.076 .
## Year2006        0.17229    0.15591    1.11   0.270
```

```

## Year2007          -0.06864      0.14273    -0.48      0.631
## Year2008           0.06364      0.14559      0.44      0.662
## Year2009           0.34193      0.19719      1.73      0.084 .
## Year2010           0.17830      0.21833      0.82      0.415
## Year2011           0.25249      0.20926      1.21      0.229
## Year2012           0.04846      0.33705      0.14      0.886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0836, Adjusted R-squared:  0.0281
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 275 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.398  0.864   0.956   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      3.34e-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: 299"
## [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]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    2   35   29   21   15   15   19   18   15   29   21   10   29   28   34
## 2012
##    33
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0   33   25   15    9   15   16   16   15   26   19    8   24   27   29
## 2012

```

```
## 30
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
## 0 33 24 14 9 15 15 16 15 24 19 7 22 26 26
## 2012
## 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 = 21, df = 14, p-value = 0.1
```



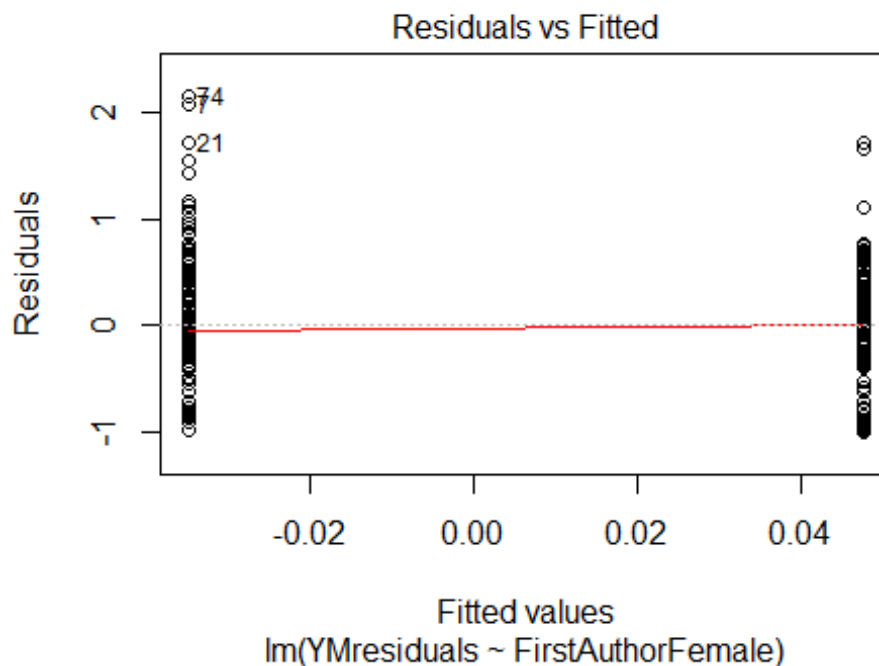
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 2.1799893566552"
## [1] "Male first author team size 2018 geometric mean: 2.76822945651829"

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

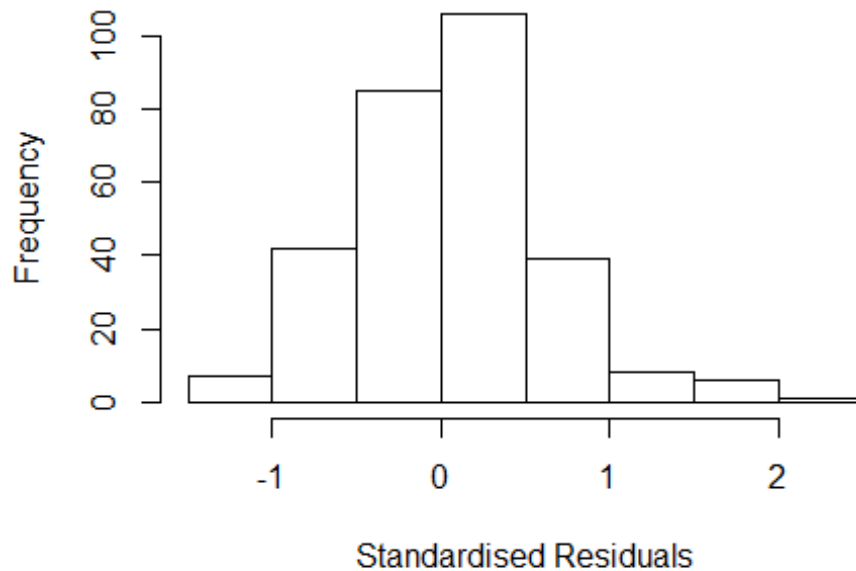
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 25, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.35296810710145"
## [1] "Male last author team size 2018 geometric mean: 2.37617679756498"

## 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 = 35, 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 1.686 1          1.299
## LastAuthorFemale 1.540 1          1.241
## UniqueAuthors    3.135 4          1.154
## Year              3.924 14         1.050
```

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
## -1.0963 -0.4185 0.0441 0.4204 2.2891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4289 0.1420 3.02 0.00276 **
## FirstAuthorFemale1 0.0702 0.0841 0.83 0.40449
## LastAuthorFemale1 -0.0239 0.0796 -0.30 0.76400
## UniqueAuthors2 0.3234 0.0933 3.47 0.00061 ***
## UniqueAuthors3 0.3429 0.0988 3.47 0.00061 ***
## UniqueAuthors4 0.2682 0.1606 1.67 0.09606 .
## UniqueAuthors5 0.1494 0.1433 1.04 0.29790
## Year1999 0.2170 0.1830 1.19 0.23687
## Year2000 0.2905 0.2936 0.99 0.32328
## Year2001 0.3803 0.2077 1.83 0.06823 .
```

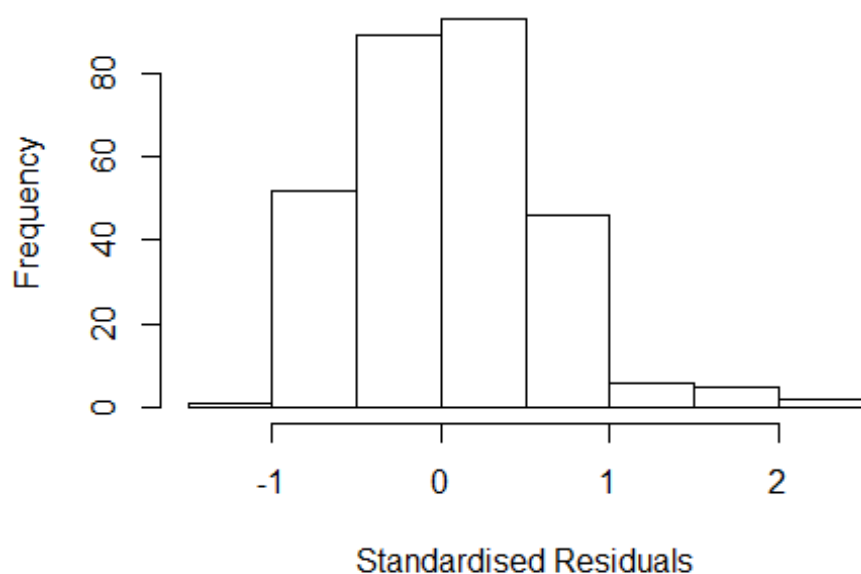


```

## Year2002          0.4352      0.2346      1.86  0.06459 .
## Year2003          0.6353      0.1872      3.39  0.00079 ***
## Year2004         -0.0104      0.1764     -0.06  0.95306
## Year2005          0.2549      0.1977      1.29  0.19836
## Year2006          0.4130      0.1614      2.56  0.01104 *
## Year2007          0.2580      0.2052      1.26  0.20981
## Year2008          0.4343      0.1842      2.36  0.01908 *
## Year2009          0.2428      0.1853      1.31  0.19129
## Year2010          0.2738      0.1796      1.52  0.12854
## Year2011          0.1311      0.1869      0.70  0.48365
## Year2012          0.2478      0.1919      1.29  0.19768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.166, Adjusted R-squared:  0.105
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 268 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0507 0.8800 0.9460 0.9020 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          3.40e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.604 1 1.267
## LastAuthorFemale 1.428 1 1.195
## Year 1.470 14 1.014

```

Residuals from first and last author



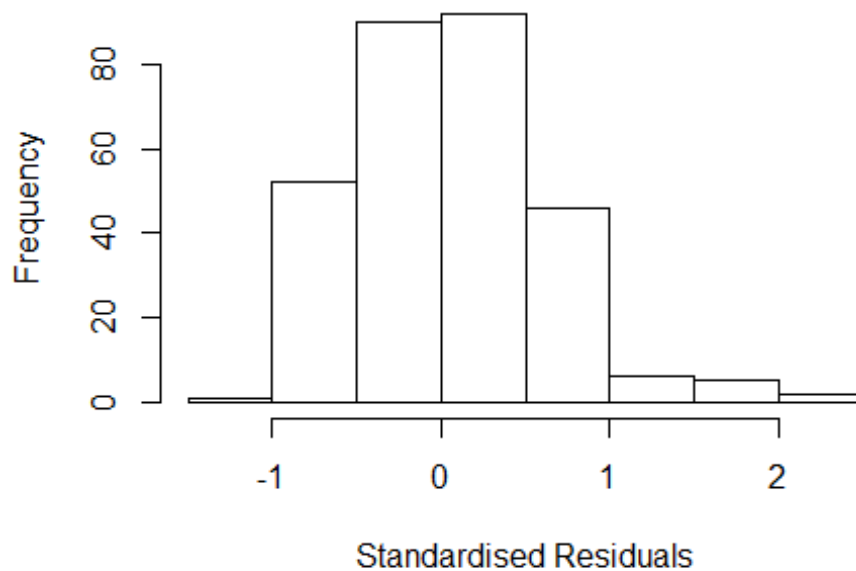
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0174 -0.4547 0.0429 0.3865 2.2413
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4767 0.1374 3.47 0.00061 ***
## FirstAuthorFemale1 0.1450 0.0861 1.68 0.09321 .
## LastAuthorFemale1 -0.0420 0.0817 -0.51 0.60780
## Year1999 0.2869 0.1953 1.47 0.14281
## Year2000 0.3313 0.2616 1.27 0.20637
## Year2001 0.4693 0.2106 2.23 0.02668 *
## Year2002 0.4967 0.2042 2.43 0.01565 *
## Year2003 0.6702 0.1751 3.83 0.00016 ***
## Year2004 0.0223 0.1857 0.12 0.90442
## Year2005 0.3538 0.2077 1.70 0.08972 .
## Year2006 0.4606 0.1666 2.76 0.00609 **
## Year2007 0.4377 0.2125 2.06 0.04040 *
```

```

## Year2008          0.5878      0.2090      2.81  0.00527 **
## Year2009          0.3391      0.1961      1.73  0.08480 .
## Year2010          0.3678      0.1809      2.03  0.04296 *
## Year2011          0.2481      0.1893      1.31  0.19112
## Year2012          0.3329      0.1891      1.76  0.07942 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.101, Adjusted R-squared:  0.049
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.114  0.878  0.947  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      3.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.299 1          1.140
## Year              1.299 14          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
## -1.0239 -0.4460 0.0363 0.3905 2.2452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4728 0.1378 3.43 0.00069 ***
## FirstAuthorFemale1 0.1212 0.0780 1.55 0.12139
## Year1999 0.2899 0.1959 1.48 0.14007
## Year2000 0.3216 0.2607 1.23 0.21829
## Year2001 0.4683 0.2105 2.22 0.02690 *
## Year2002 0.4978 0.2056 2.42 0.01610 *
## Year2003 0.6682 0.1737 3.85 0.00015 ***
## Year2004 0.0223 0.1855 0.12 0.90448
## Year2005 0.3425 0.2028 1.69 0.09243 .
## Year2006 0.4640 0.1674 2.77 0.00595 **
## Year2007 0.4298 0.2099 2.05 0.04154 *
## Year2008 0.5810 0.2057 2.82 0.00508 **
```

```

## Year2009          0.3349      0.1966      1.70  0.08956 .
## Year2010          0.3719      0.1813      2.05  0.04115 *
## Year2011          0.2449      0.1906      1.29  0.19980
## Year2012          0.3325      0.1894      1.76  0.08030 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0521
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.878  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      3.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.18  1          1.086
## Year              1.18 14          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
## -1.0058 -0.4663  0.0112  0.3870  2.2241
##
## Coefficients:

```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4939    0.1366   3.62  0.00036 ***
## LastAuthorFemale1 0.0376    0.0751   0.50  0.61721
## Year1999          0.2839    0.1957   1.45  0.14802
## Year2000          0.2989    0.2566   1.17  0.24501
## Year2001          0.4987    0.2184   2.28  0.02315 *
## Year2002          0.5119    0.2008   2.55  0.01135 *
## Year2003          0.6811    0.1736   3.92  0.00011 ***
## Year2004          0.0174    0.1879   0.09  0.92623
## Year2005          0.3411    0.2049   1.66  0.09709 .
## Year2006          0.5093    0.1666   3.06  0.00245 **
## Year2007          0.4588    0.2130   2.15  0.03206 *
## Year2008          0.5792    0.2074   2.79  0.00560 **
## Year2009          0.3597    0.1933   1.86  0.06383 .
## Year2010          0.4138    0.1734   2.39  0.01768 *
## Year2011          0.2715    0.1877   1.45  0.14906
## Year2012          0.3751    0.1886   1.99  0.04768 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.6
## Multiple R-squared:  0.0911, Adjusted R-squared:  0.0421
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 272 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.140  0.867  0.946  0.911  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      3.40e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 294"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3602"
## [1] "#####"

```

```

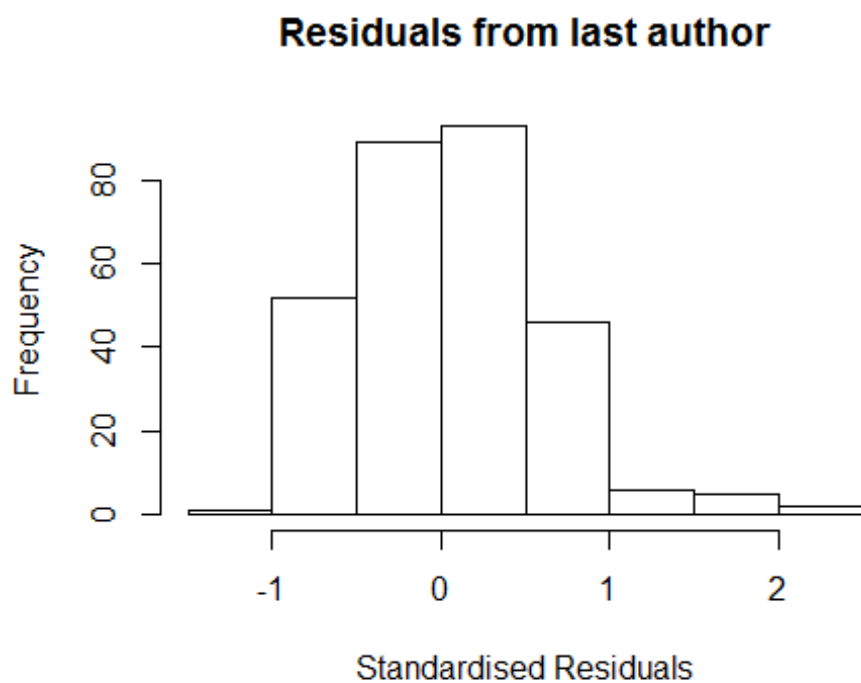
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    4    4    6   11    6    9    7    2   15    5    8    8    6    4    6
## 2012
##    10
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    1    0    7    4    8    4    1   13    5    6    8    6    4    6
## 2012
##    10
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    0    1    0    6    4    8    4    1   12    5    4    8    6    4    6
## 2012
##    10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.26843458790854"
## [1] "Male first author team size 2018 geometric mean: 1.91054600869993"

## 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 = 50, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.02190013558564"
## [1] "Male last author team size 2018 geometric mean: 2.40475854740257"

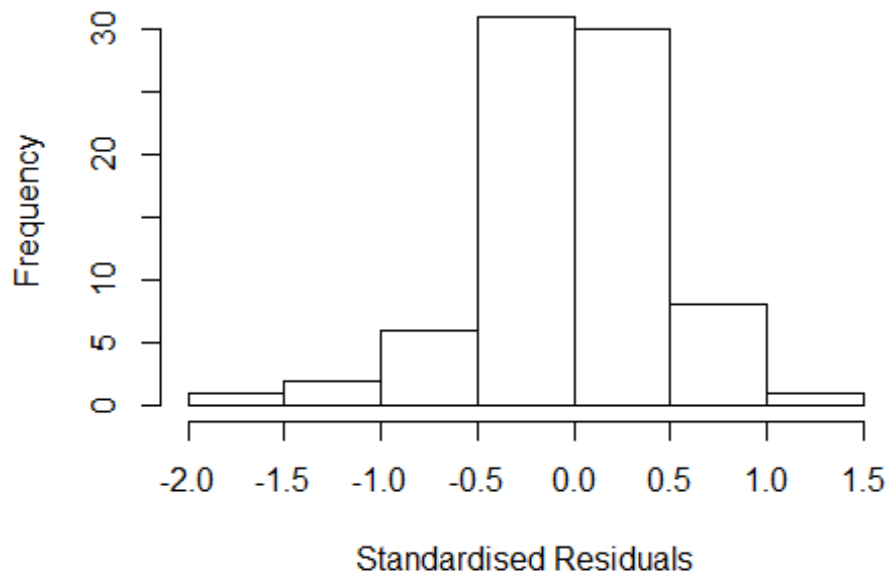
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32, 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 -6.630e+14  1      NaN
## LastAuthorFemale  -1.170e+15  1      NaN
## UniqueAuthors    -1.746e+17  4      NaN
## Year              -2.971e+17 13      NaN
```


Residuals from first and last author and team size



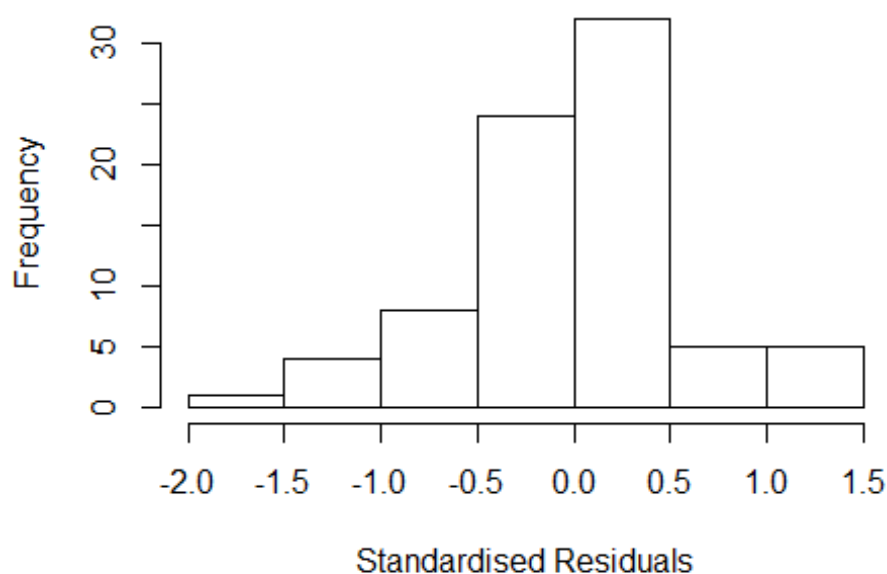
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.56e+00 -2.48e-01 2.22e-16 2.20e-01 1.22e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3300 0.1967 1.68 0.09869 .
## FirstAuthorFemale1 -0.0823 0.1385 -0.59 0.55470
## LastAuthorFemale1 -0.2477 0.1839 -1.35 0.18311
## UniqueAuthors2 0.4745 0.2190 2.17 0.03427 *
## UniqueAuthors3 0.5885 0.2039 2.89 0.00544 **
## UniqueAuthors4 1.3267 0.1777 7.46 4.4e-10 ***
## UniqueAuthors5 0.6046 0.2013 3.00 0.00392 **
## Year2000 0.6336 0.1694 3.74 0.00042 ***
## Year2001 0.8722 0.1695 5.15 3.2e-06 ***
## Year2002 0.6672 0.1422 4.69 1.7e-05 ***
```

```

## Year2003          0.6944      0.1574      4.41  4.4e-05 ***
## Year2004          0.7255      0.1699      4.27  7.2e-05 ***
## Year2005          0.6747      0.1995      3.38  0.00128 **
## Year2006          0.4231      0.2945      1.44  0.15606
## Year2007          0.6224      0.2784      2.24  0.02917 *
## Year2008          0.3065      0.2378      1.29  0.20235
## Year2009          0.3401      0.2346      1.45  0.15252
## Year2010         -0.0407      0.1466     -0.28  0.78241
## Year2011          0.4875      0.2332      2.09  0.04087 *
## Year2012          0.3663      0.2178      1.68  0.09792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.447
## Multiple R-squared:  0.485, Adjusted R-squared:  0.319
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
##  11 weights are ~ = 1. The remaining 68 ones are summarized as
##    Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    0.199  0.873   0.966   0.901   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      1.27e-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 1.992e+14 1      1.412e+07
## LastAuthorFemale  1.384e+14 1      1.176e+07
## Year              1.031e+15 13      3.779e+00

```

Residuals from first and last author



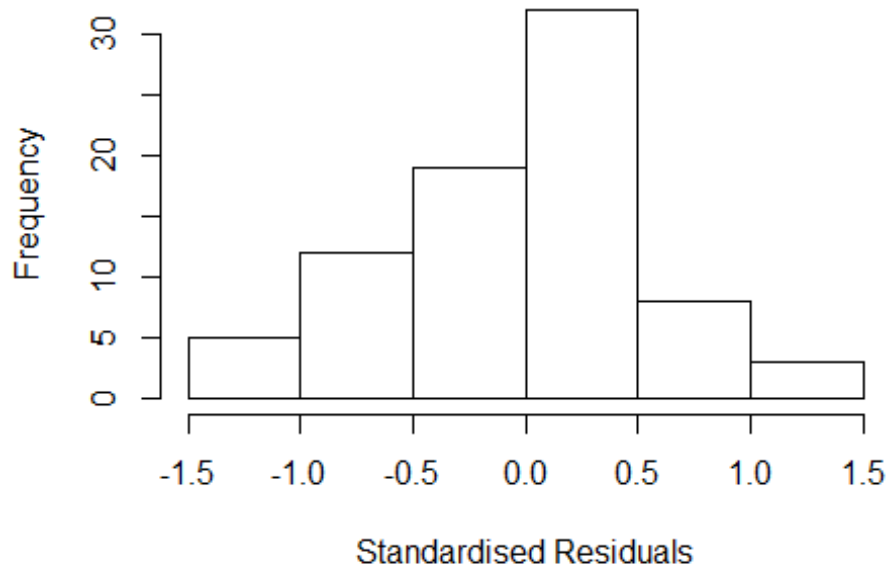
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5165 -0.3273 0.0288 0.2895 1.3845
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.533 0.268 1.99 0.05147 .
## FirstAuthorFemale1 -0.122 0.308 -0.40 0.69209
## LastAuthorFemale1 -0.410 0.256 -1.60 0.11444
## Year2000 0.683 0.226 3.02 0.00366 **
## Year2001 1.014 0.272 3.73 0.00041 ***
## Year2002 0.762 0.210 3.63 0.00056 ***
## Year2003 1.002 0.186 5.40 1.1e-06 ***
## Year2004 0.997 0.268 3.72 0.00043 ***
## Year2005 0.855 0.232 3.69 0.00048 ***
## Year2006 0.804 0.511 1.57 0.12030
## Year2007 0.984 0.321 3.06 0.00323 **
## Year2008 0.583 0.479 1.22 0.22805
```

```

## Year2009          0.709      0.433      1.64  0.10635
## Year2010          0.467      0.408      1.15  0.25650
## Year2011          0.735      0.189      3.88  0.00025 ***
## Year2012          0.595      0.379      1.57  0.12163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.24,   Adjusted R-squared:  0.0585
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 70 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.268  0.874  0.949  0.871  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      1.27e-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 5.629e+14 1      2.373e+07
## Year              5.629e+14 13      3.693e+00

```

Residuals from first author



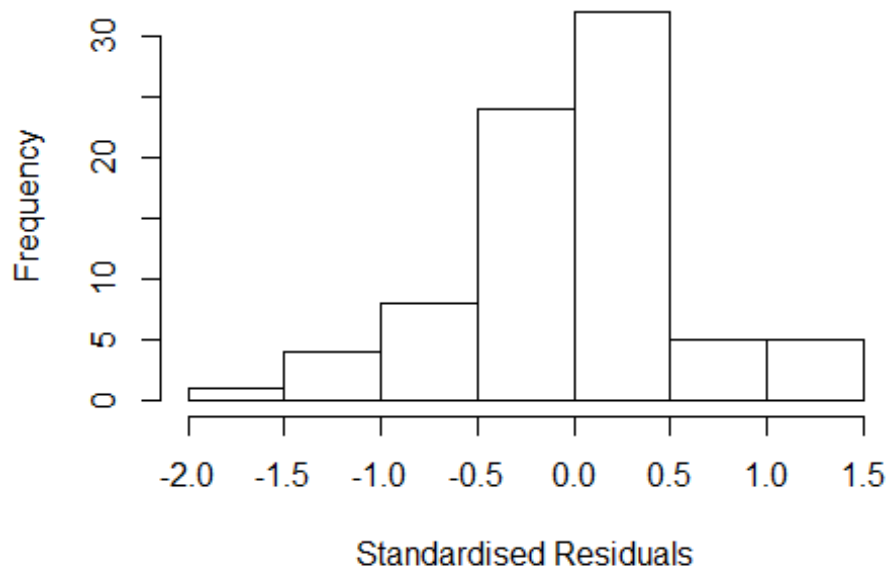
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.408 -0.416 0.044 0.319 1.448
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.312 0.215 1.45 0.1509
## FirstAuthorFemale1 -0.312 0.215 -1.45 0.1509
## Year2000 0.832 0.211 3.95 0.0002 ***
## Year2001 1.158 0.376 3.08 0.0030 **
## Year2002 0.900 0.208 4.33 5.4e-05 ***
## Year2003 1.268 0.136 9.30 1.7e-13 ***
## Year2004 1.218 0.215 5.67 3.7e-07 ***
## Year2005 0.990 0.238 4.16 9.7e-05 ***
## Year2006 1.011 0.394 2.57 0.0127 *
## Year2007 1.095 0.427 2.56 0.0127 *
## Year2008 0.777 0.327 2.38 0.0205 *
## Year2009 0.838 0.389 2.16 0.0349 *
```

```

## Year2010          0.626      0.267      2.34    0.0223 *
## Year2011          0.985      0.127      7.78    7.9e-11 ***
## Year2012          0.744      0.363      2.05    0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.143, Adjusted R-squared:  -0.0449
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 67 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.487  0.886  0.951  0.900  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      1.27e-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 2.252e+14  1      1.501e+07
## Year              2.252e+14 13      3.565e+00

```

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
## -1.4361 -0.3471 0.0374 0.3255 1.4157
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.445 0.187 2.38 0.02027 *
## LastAuthorFemale1 -0.445 0.187 -2.38 0.02027 *
## Year2000 0.742 0.185 4.02 0.00016 ***
## Year2001 1.033 0.255 4.05 0.00014 ***
## Year2002 0.815 0.181 4.51 2.8e-05 ***
## Year2003 1.005 0.174 5.77 2.5e-07 ***
## Year2004 1.085 0.187 5.80 2.2e-07 ***
## Year2005 0.905 0.190 4.78 1.1e-05 ***
## Year2006 0.895 0.488 1.83 0.07127 .
## Year2007 0.991 0.403 2.46 0.01659 *
## Year2008 0.603 0.424 1.42 0.15977
## Year2009 0.800 0.320 2.50 0.01492 *
```

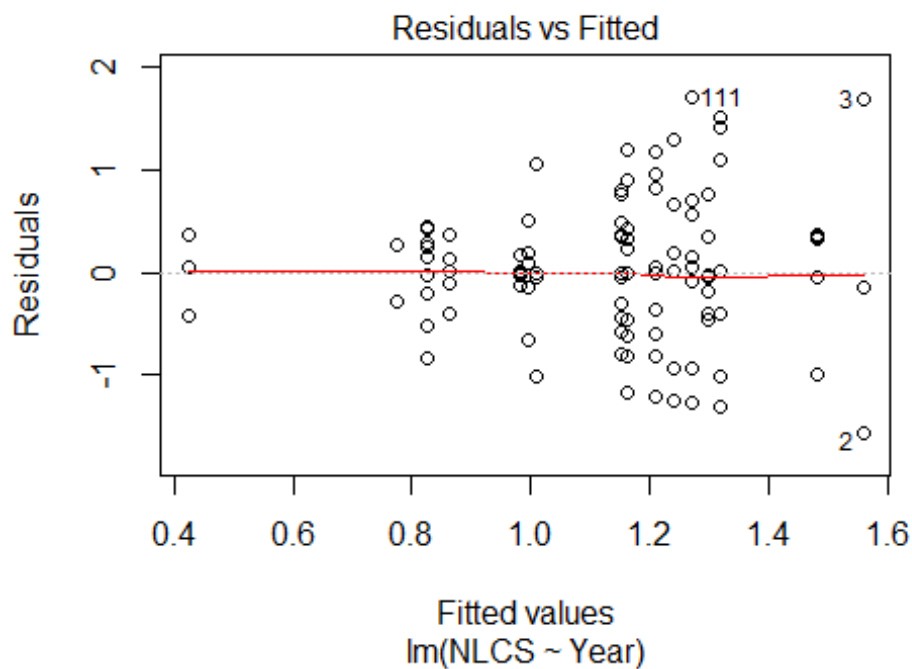
```

## Year2010          0.571      0.342      1.67  0.09958 .
## Year2011          0.767      0.177      4.34  5.2e-05 ***
## Year2012          0.684      0.273      2.51  0.01477 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.53
## Multiple R-squared:  0.199, Adjusted R-squared:  0.0239
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 70 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.443  0.897  0.958  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      1.27e-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: 79"
## [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
##    3    2    3   13    6   13    8    7   10    5    7    8    8   12   12
## 2011 2012
##   15   13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3    2    3    9    0    4    5    5    8    5    5    6    7   10   10
## 2011 2012
##   12    9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```



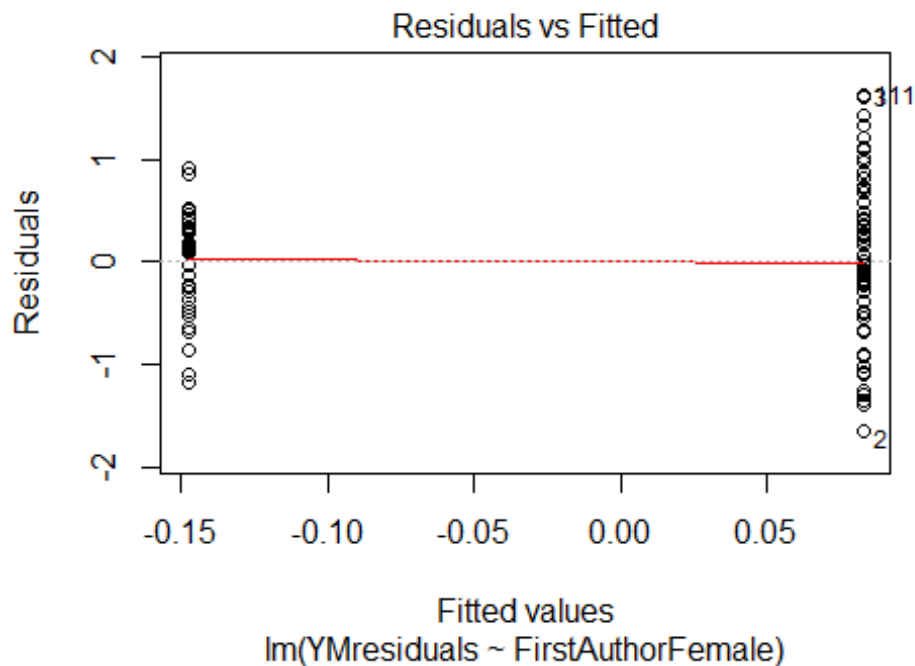
```
##      3      2      3      8      0      4      4      2      8      5      5      6      7      9      10
## 2011 2012
##    11     8
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 35, df = 15, p-value = 0.003
```



```
##
## 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: 2.44948974278318"
## [1] "Male first author team size 2018 geometric mean: 6"
##
## 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 = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77976314968462"
## [1] "Male last author team size 2018 geometric mean: 2.88449914061482"

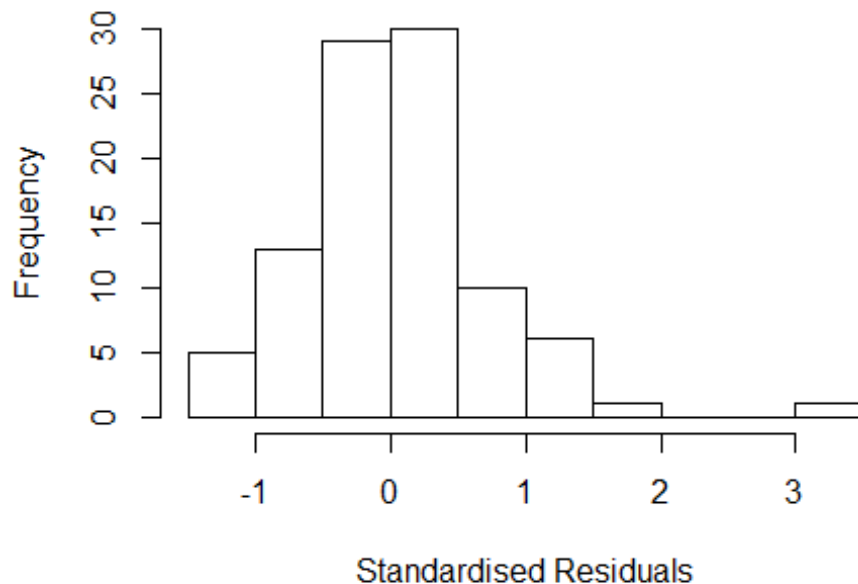
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6.5, 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	12.210	1	3.494
LastAuthorFemale	5.858	1	2.420
UniqueAuthors	27.659	4	1.514
Year	423.377	15	1.223

Residuals from first and last author and team size



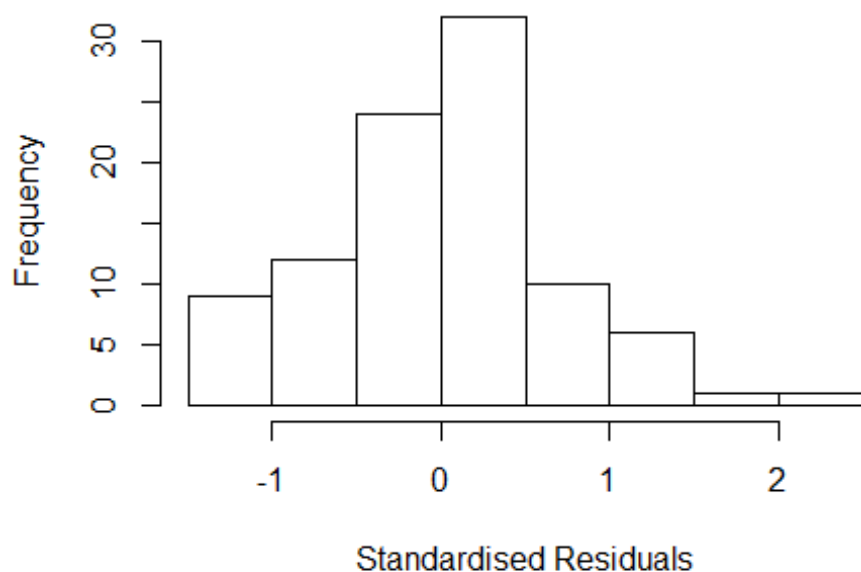
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3 0009035705 3.26 1996      1210      2      3.094
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3649 -0.3725  0.0244  0.3279  3.0942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.068      0.632   0.11  0.91465
## FirstAuthorFemale1  0.217      0.165   1.32  0.19094
## LastAuthorFemale1 -0.119      0.169  -0.71  0.48212
## UniqueAuthors2     0.484      0.187   2.59  0.01157 *
## UniqueAuthors3     0.380      0.214   1.78  0.07984 .
## UniqueAuthors4     0.725      0.234   3.09  0.00280 **
## UniqueAuthors5     0.820      0.235   3.49  0.00082 ***
## Year1997           0.464      0.607   0.76  0.44705
## Year1998          -0.110      0.586  -0.19  0.85116
## Year1999           0.387      0.582   0.67  0.50802
```

```

## Year2001          0.684      0.797      0.86  0.39338
## Year2002          0.479      0.612      0.78  0.43591
## Year2003          0.247      0.695      0.35  0.72374
## Year2004          0.781      0.799      0.98  0.33126
## Year2005          0.533      0.621      0.86  0.39284
## Year2006          0.478      0.618      0.77  0.44120
## Year2007          0.602      0.703      0.86  0.39486
## Year2008          0.639      0.622      1.03  0.30789
## Year2009          0.421      0.691      0.61  0.54438
## Year2010          0.595      0.643      0.93  0.35754
## Year2011          0.524      0.624      0.84  0.40416
## Year2012          0.596      0.641      0.93  0.35570
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.224, Adjusted R-squared:  0.00139
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 2 is an outlier with |weight| = 0 ( < 0.0011);
## 10 weights are ~= 1. The remaining 84 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.376  0.875   0.963   0.914   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          1.05e-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 2.218  1         1.489
## LastAuthorFemale  4.007  1         2.002
## Year              8.155 15         1.072

```

Residuals from first and last author



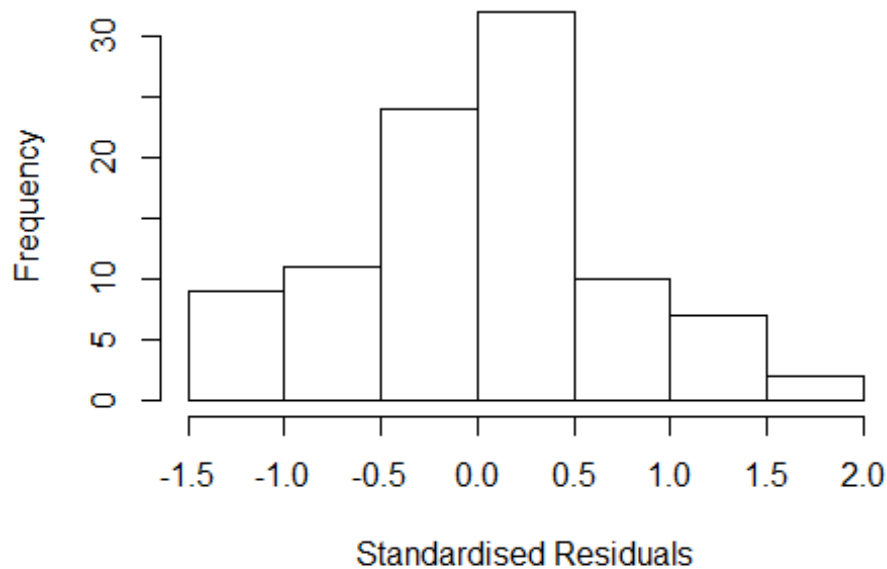
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3467 -0.4115 0.0265 0.3320 2.0911
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0112 1.6705 0.61 0.55
## FirstAuthorFemale1 0.2691 0.1673 1.61 0.11
## LastAuthorFemale1 -0.1114 0.2223 -0.50 0.62
## Year1997 -0.2372 1.6823 -0.14 0.89
## Year1998 -0.8194 1.6921 -0.48 0.63
## Year1999 -0.2497 1.7194 -0.15 0.88
## Year2001 -0.0763 1.8461 -0.04 0.97
## Year2002 -0.1838 1.7689 -0.10 0.92
## Year2003 -0.1266 1.8319 -0.07 0.95
## Year2004 0.0959 1.7351 0.06 0.96
## Year2005 -0.1575 1.7215 -0.09 0.93
## Year2006 -0.0707 1.7457 -0.04 0.97
```

```

## Year2007          0.1914      1.7816      0.11      0.91
## Year2008          0.1288      1.7492      0.07      0.94
## Year2009         -0.0134      1.7838     -0.01      0.99
## Year2010          0.0664      1.7890      0.04      0.97
## Year2011         -0.0118      1.7158     -0.01      0.99
## Year2012          0.0434      1.7299      0.03      0.98
##
## Robust residual standard error: 0.707
## Multiple R-squared:  0.0933, Adjusted R-squared:  -0.107
## Convergence in 48 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 83 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.362  0.851  0.966  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      1.05e-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 1.941 1          1.393
## Year              1.941 15          1.022

```

Residuals from first author



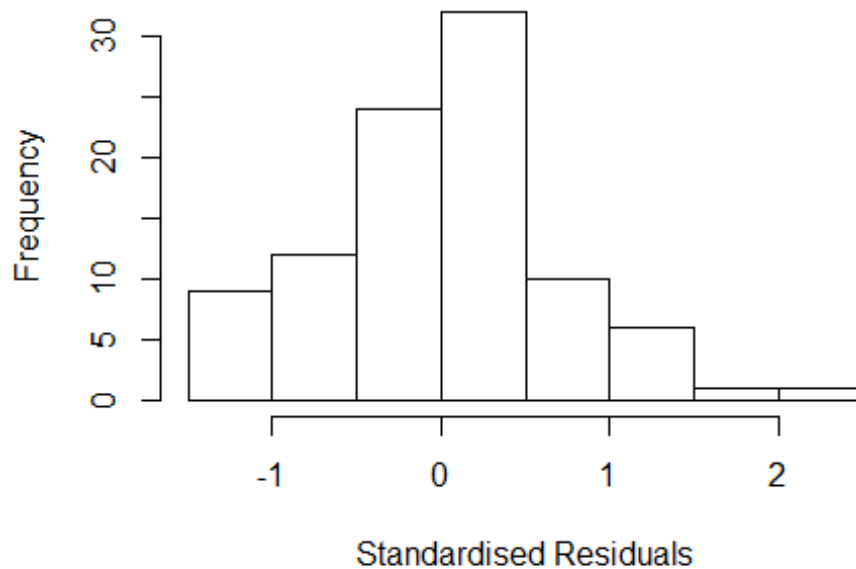
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3874 -0.4236 0.0581 0.3577 1.8726
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13490 1.58600 0.72 0.48
## FirstAuthorFemale1 0.25255 0.16228 1.56 0.12
## Year1997 -0.36090 1.59840 -0.23 0.82
## Year1998 -0.96255 1.59138 -0.60 0.55
## Year1999 -0.41522 1.58897 -0.26 0.79
## Year2001 -0.30213 1.63585 -0.18 0.85
## Year2002 -0.37951 1.58808 -0.24 0.81
## Year2003 -0.29767 1.64407 -0.18 0.86
## Year2004 -0.01800 1.66932 -0.01 0.99
## Year2005 -0.32199 1.58918 -0.20 0.84
## Year2006 -0.25486 1.58412 -0.16 0.87
## Year2007 0.00998 1.64133 0.01 1.00
```

```

## Year2008          -0.05946    1.58500   -0.04    0.97
## Year2009          -0.17672    1.61727   -0.11    0.91
## Year2010          -0.14198    1.59842   -0.09    0.93
## Year2011          -0.17702    1.58541   -0.11    0.91
## Year2012          -0.12598    1.61231   -0.08    0.94
##
## Robust residual standard error: 0.746
## Multiple R-squared: 0.091, Adjusted R-squared: -0.0955
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 84 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.509  0.867  0.970  0.916  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      1.05e-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 3.231 1          1.798
## Year            3.231 15          1.040

```


Residuals from last author



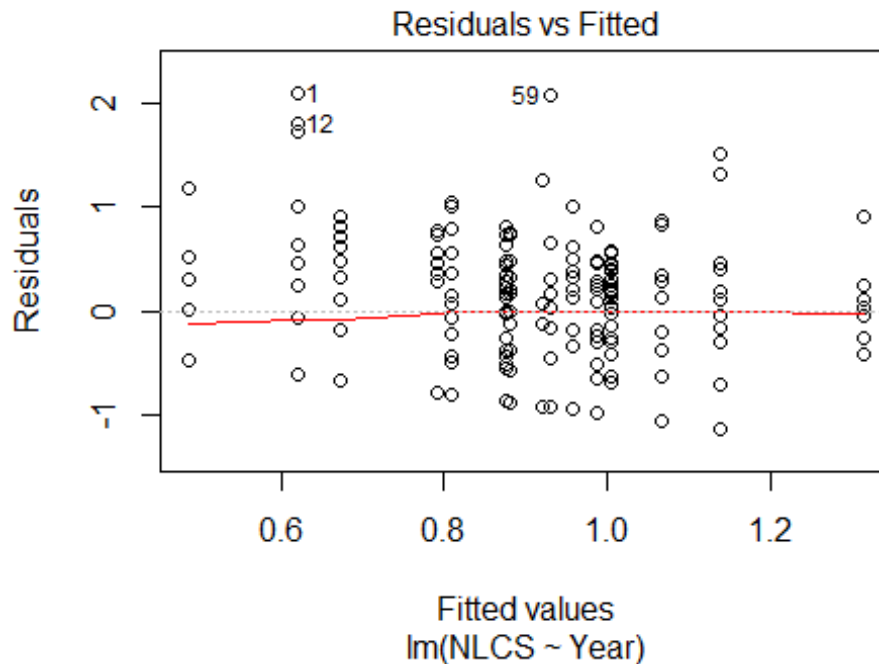
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28480 -0.42281 0.00961 0.34328 1.97520
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3466 1.6128 0.83 0.41
## LastAuthorFemale1 -0.0618 0.2034 -0.30 0.76
## Year1997 -0.5726 1.6250 -0.35 0.73
## Year1998 -0.9018 1.6604 -0.54 0.59
## Year1999 -0.4116 1.6790 -0.25 0.81
## Year2001 -0.2834 1.7850 -0.16 0.87
## Year2002 -0.3493 1.7335 -0.20 0.84
## Year2003 -0.3522 1.7288 -0.20 0.84
## Year2004 -0.1097 1.7351 -0.06 0.95
## Year2005 -0.4580 1.6640 -0.28 0.78
## Year2006 -0.3288 1.6873 -0.19 0.85
## Year2007 -0.0603 1.7304 -0.03 0.97
```

```

## Year2008          -0.0140      1.7045   -0.01      0.99
## Year2009          -0.1753      1.7179   -0.10      0.92
## Year2010          -0.1339      1.7349   -0.08      0.94
## Year2011          -0.2409      1.6757   -0.14      0.89
## Year2012          -0.0504      1.7295   -0.03      0.98
##
## Robust residual standard error: 0.728
## Multiple R-squared:  0.0689, Adjusted R-squared:  -0.122
## Convergence in 42 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 77 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.442  0.863  0.957  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           1.05e-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: 95"
## [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 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   33   23   18    9   12   10   17   22   39   26   20   29   18   42   53
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   31   20   14    6   12    9   15   20   29   14   10   22   17   13   20
##
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##   31   19   14    6   12    9   15   19   27   12    9   20   16   11   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 = 25, df = 14, p-value = 0.03
```



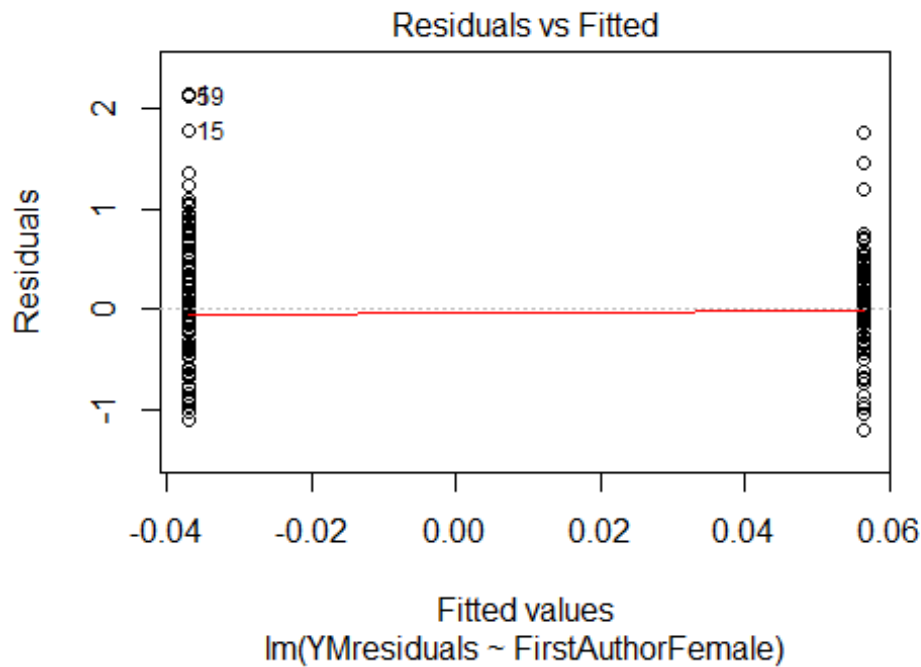
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.4, df = 1, p-value = 0.007

## [1] "Female first author team size 2018 geometric mean: 3.60738064640132"
## [1] "Male first author team size 2018 geometric mean: 2.65160197599675"

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

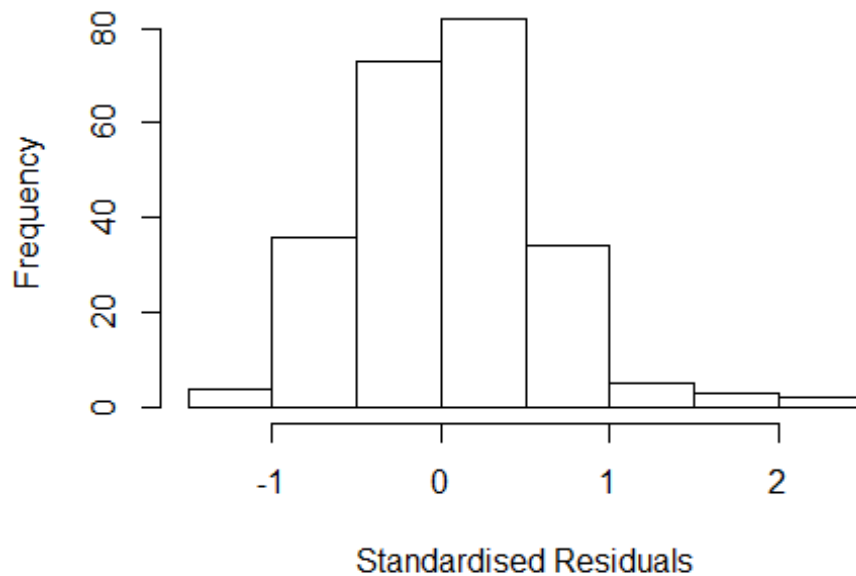
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.42485189209207"
## [1] "Male last author team size 2018 geometric mean: 3.02422922216055"
```

```
## 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 3.000 1          1.732
## LastAuthorFemale  2.880 1          1.697
## UniqueAuthors    3.968 4          1.188
## Year              4.934 14         1.059
```

Residuals from first and last author and team size



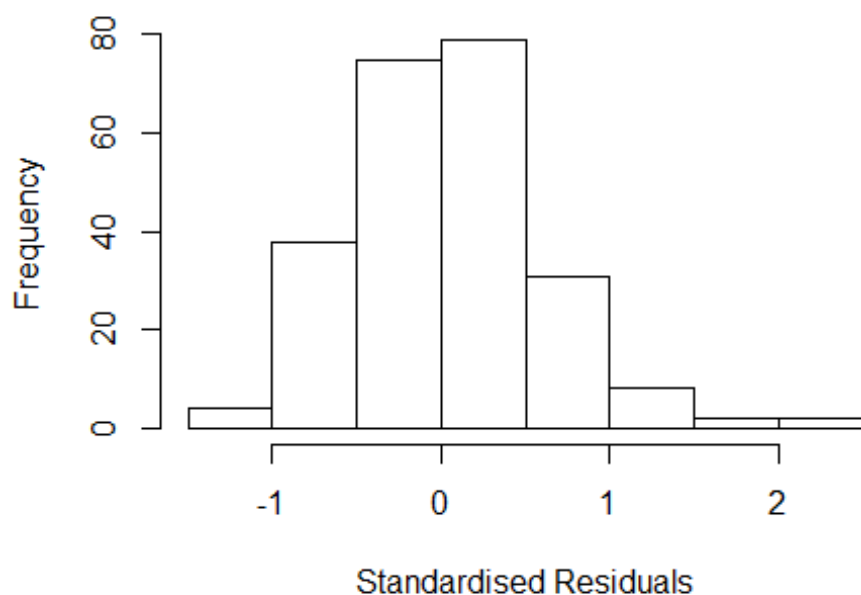
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1172 -0.3751 0.0285 0.3854 2.3429
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.375144 0.135408 2.77 0.0061 **
## FirstAuthorFemale1 0.033271 0.120158 0.28 0.7821
## LastAuthorFemale1 0.093751 0.111540 0.84 0.4015
## UniqueAuthors2 0.276083 0.104782 2.63 0.0090 **
## UniqueAuthors3 0.309319 0.169935 1.82 0.0701 .
## UniqueAuthors4 0.176090 0.163124 1.08 0.2816
## UniqueAuthors5 0.109850 0.153470 0.72 0.4749
## Year1999 0.220370 0.193819 1.14 0.2568
## Year2000 0.313903 0.298716 1.05 0.2945
## Year2001 0.384912 0.267204 1.44 0.1512
```

```

## Year2002      0.587634    0.242762    2.42    0.0163 *
## Year2003      0.868672    0.177660    4.89    2e-06 ***
## Year2004      0.000594    0.170868    0.00    0.9972
## Year2005      0.400229    0.185109    2.16    0.0317 *
## Year2006      0.436895    0.162036    2.70    0.0076 **
## Year2007      0.225356    0.218434    1.03    0.3034
## Year2008      0.122976    0.298342    0.41    0.6806
## Year2009      0.356295    0.184553    1.93    0.0548 .
## Year2010      0.280604    0.197704    1.42    0.1572
## Year2011      0.318657    0.216325    1.47    0.1422
## Year2012      0.532647    0.212947    2.50    0.0131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.191, Adjusted R-squared:  0.116
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0315 0.8740 0.9530 0.8990 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      4.18e-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 2.314 1 1.521
## LastAuthorFemale 1.877 1 1.370
## Year 2.206 14 1.029

```

Residuals from first and last author



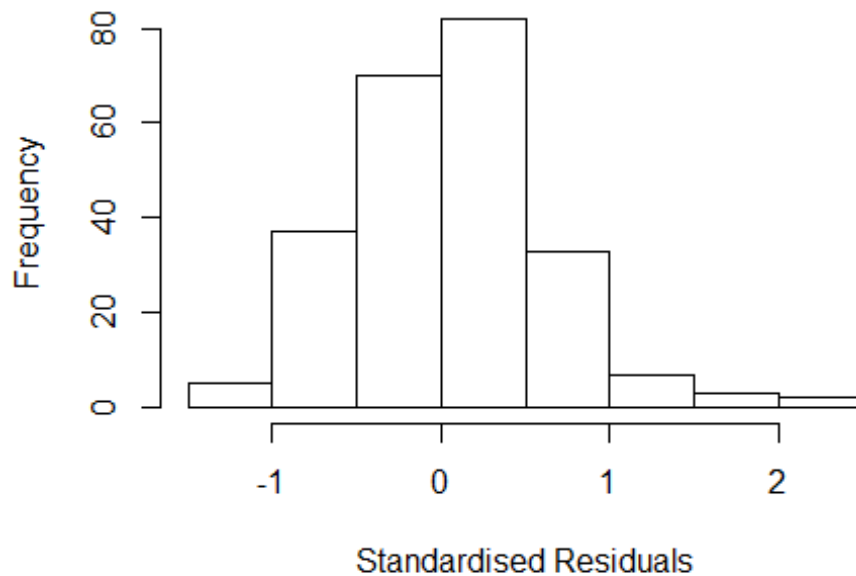
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0853 -0.4078 0.0122 0.4202 2.3102
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4078 0.1325 3.08 0.0023 **
## FirstAuthorFemale1 0.0743 0.1081 0.69 0.4923
## LastAuthorFemale1 0.0920 0.0963 0.96 0.3403
## Year1999 0.2689 0.2052 1.31 0.1915
## Year2000 0.3525 0.2654 1.33 0.1854
## Year2001 0.4102 0.2604 1.58 0.1166
## Year2002 0.6254 0.2248 2.78 0.0059 **
## Year2003 0.8235 0.1774 4.64 5.9e-06 ***
## Year2004 0.0265 0.1777 0.15 0.8817
## Year2005 0.4906 0.1893 2.59 0.0102 *
## Year2006 0.4833 0.1640 2.95 0.0036 **
## Year2007 0.3785 0.2345 1.61 0.1079
```

```

## Year2008          0.2530      0.3155      0.80      0.4234
## Year2009          0.4308      0.1910      2.26      0.0251 *
## Year2010          0.3125      0.2117      1.48      0.1414
## Year2011          0.4012      0.2233      1.80      0.0738 .
## Year2012          0.6032      0.2152      2.80      0.0055 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.15,   Adjusted R-squared:  0.089
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0629 0.8750 0.9480 0.9020 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      4.18e-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 1.754 1          1.324
## Year              1.754 14          1.020

```


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
## -1.1312 -0.4097 0.0439 0.4250 2.3083
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4097 0.1314 3.12 0.0021 **
## FirstAuthorFemale1 0.1210 0.0941 1.29 0.1996
## Year1999 0.2687 0.2170 1.24 0.2170
## Year2000 0.3764 0.2592 1.45 0.1479
## Year2001 0.4164 0.2624 1.59 0.1139
## Year2002 0.6251 0.2207 2.83 0.0050 **
## Year2003 0.8372 0.1753 4.78 3.2e-06 ***
## Year2004 0.0316 0.1759 0.18 0.8574
## Year2005 0.5214 0.1878 2.78 0.0060 **
## Year2006 0.4957 0.1606 3.09 0.0023 **
## Year2007 0.3946 0.2356 1.67 0.0954 .
## Year2008 0.2881 0.3297 0.87 0.3832
```

```

## Year2009          0.4404      0.1879      2.34      0.0199 *
## Year2010          0.3224      0.2120      1.52      0.1297
## Year2011          0.4090      0.2218      1.84      0.0664 .
## Year2012          0.6005      0.2137      2.81      0.0054 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.15,   Adjusted R-squared:  0.0924
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0521 0.8640 0.9440 0.8980 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      4.18e-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 1.444 1          1.202
## Year            1.444 14          1.013

## [1] "List of 0 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
## -1.0546 -0.4088  0.0234  0.4114  2.3092
##
## Coefficients:

```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4088    0.1322   3.09  0.0022 **
## LastAuthorFemale1 0.1269    0.0838   1.51  0.1314
## Year1999          0.2693    0.2179   1.24  0.2177
## Year2000           0.3403    0.2623   1.30  0.1958
## Year2001           0.4160    0.2622   1.59  0.1140
## Year2002           0.6421    0.2210   2.91  0.0040 **
## Year2003           0.8359    0.1753   4.77  3.4e-06 ***
## Year2004           0.0294    0.1778   0.17  0.8687
## Year2005           0.5035    0.1857   2.71  0.0072 **
## Year2006           0.5151    0.1586   3.25  0.0013 **
## Year2007           0.4041    0.2275   1.78  0.0770 .
## Year2008           0.2705    0.3209   0.84  0.4002
## Year2009           0.4608    0.1847   2.49  0.0133 *
## Year2010           0.3347    0.2041   1.64  0.1024
## Year2011           0.4086    0.2222   1.84  0.0673 .
## Year2012           0.6457    0.2022   3.19  0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.151, Adjusted R-squared:  0.0941
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 219 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0519 0.8640 0.9470 0.9000 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           4.18e-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: 239"
## [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]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    4    1    1    1    5    4    4    9    6    4    9   10    6    9    6
## 2012
##    12
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    0    0    1    2    3    3    4    4    3    6    8    5    7    6
## 2012
##     7
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    0    0    0    1    3    2    3    3    2    4    7    5    6    4
## 2012
##     7
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4.62666900108367"
## [1] "Male first author team size 2018 geometric mean: 3.14752698625914"

## 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 = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.32707863238384"
## [1] "Male last author team size 2018 geometric mean: 3.81071720945362"

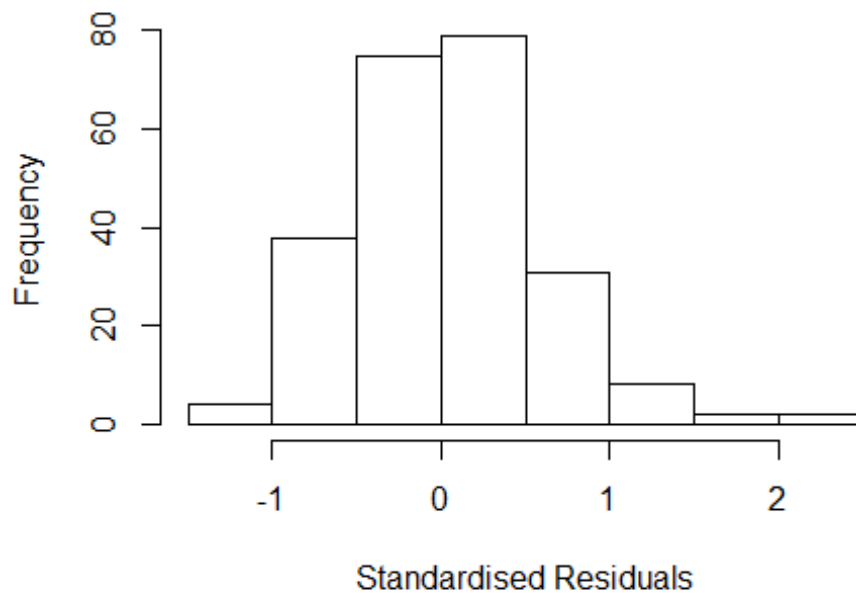
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## 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"

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

```

Residuals from last author



```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 9.772e+00  1      3.126
## LastAuthorFemale 4.571e+00  1      2.138
## UniqueAuthors    2.802e+15  4     85.297
## Year             4.611e+16 12      4.947

## [1] "List of 0 outliers with 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
## -1.2843 -0.2183  0.0227  0.2145  0.6841
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90600    0.00000      Inf < 2e-16 ***
## FirstAuthorFemale1 -0.16318    0.16468   -0.99   0.330
## LastAuthorFemale1  0.22200    0.14546    1.53   0.138
## UniqueAuthors2    0.33934    0.25678    1.32   0.197
```

```

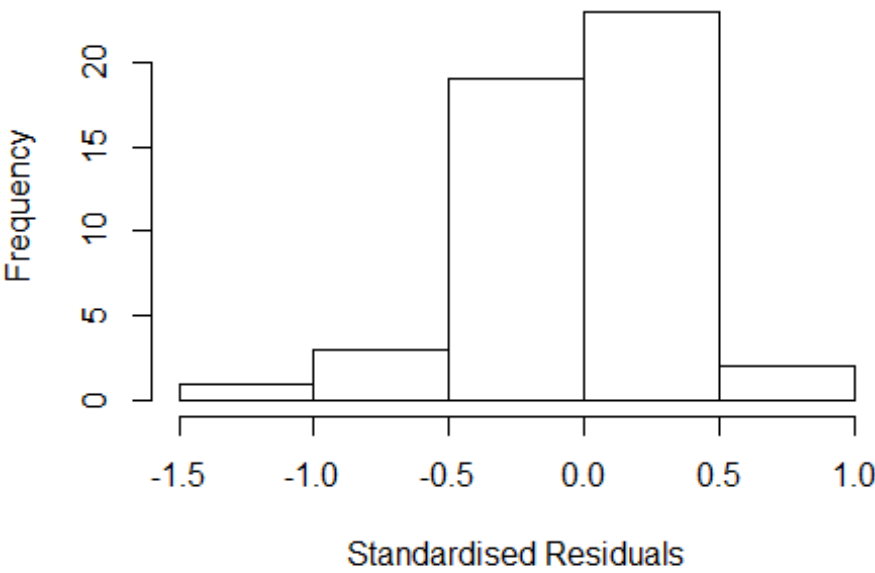
## UniqueAuthors3      0.11692      0.13802      0.85      0.404
## UniqueAuthors4      0.80082      0.29733      2.69      0.012 *
## UniqueAuthors5      0.28489      0.16966      1.68      0.104
## Year2001             -0.54282      0.29733     -1.83      0.078 .
## Year2002             -0.69080      0.25907     -2.67      0.012 *
## Year2003             -0.23632      0.55108     -0.43      0.671
## Year2004              0.00765      0.20097      0.04      0.970
## Year2005             -0.48849      0.36565     -1.34      0.192
## Year2006             -0.76300      0.10377     -7.35      4.2e-08 ***
## Year2007             -0.51923      0.49682     -1.05      0.305
## Year2008              0.20210      0.22102      0.91      0.368
## Year2009             -0.41210      0.32035     -1.29      0.208
## Year2010              0.10116      0.18577      0.54      0.590
## Year2011              0.28106      0.19086      1.47      0.152
## Year2012              0.14373      0.20415      0.70      0.487
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.54,   Adjusted R-squared:  0.255
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 44 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.237  0.896  0.968  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      2.08e-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"

## 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

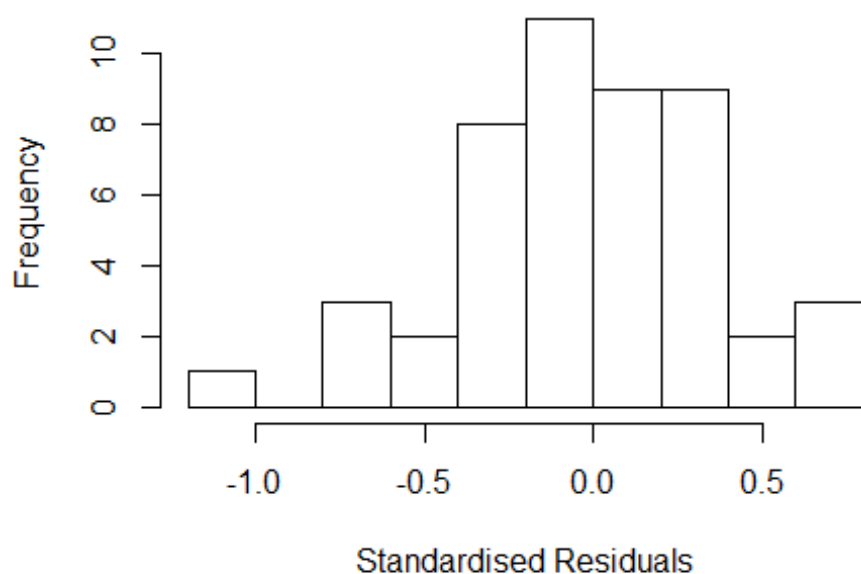
```

Residuals from first and last author and team size



##		GVIF	Df	GVIF^(1/(2*Df))
##	FirstAuthorFemale	NaN	1	NaN
##	LastAuthorFemale	NaN	1	NaN
##	Year	NaN	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
## -1.18e+00 -2.62e-01 1.03e-15 2.52e-01 6.50e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.06e-01 1.77e-08 5.12e+07 < 2e-16 ***
## FirstAuthorFemale1 -6.35e-02 1.62e-01 -3.90e-01 0.6973
## LastAuthorFemale1 1.70e-01 2.12e-01 8.00e-01 0.4292
## Year2001 2.58e-01 0.00e+00 Inf < 2e-16 ***
## Year2002 -4.52e-01 2.64e-01 -1.72e+00 0.0955 .
## Year2003 5.65e-01 4.39e-01 1.29e+00 0.2074
## Year2004 1.28e-02 1.53e-01 8.00e-02 0.9334
## Year2005 -3.82e-02 3.43e-01 -1.10e-01 0.9122
## Year2006 -7.63e-01 1.03e-01 -7.40e+00 1.7e-08 ***
## Year2007 1.02e-01 3.19e-01 3.20e-01 0.7501
## Year2008 3.40e-01 2.28e-01 1.49e+00 0.1454
## Year2009 -4.65e-02 2.15e-01 -2.20e-01 0.8303
```



```

## Year2010          2.44e-01  1.23e-01  1.99e+00  0.0548 .
## Year2011          4.22e-01  1.25e-01  3.39e+00  0.0018 **
## Year2012          5.27e-01  1.71e-01  3.07e+00  0.0042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.41,   Adjusted R-squared:  0.16
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 44 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.448  0.926  0.959  0.927  0.988  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      2.08e-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 3.758 1          1.939
## Year              3.758 12          1.057

## [1] "List of 0 outliers 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
## -1.2479 -0.2678  0.0145  0.3055  0.7782
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)

```

```

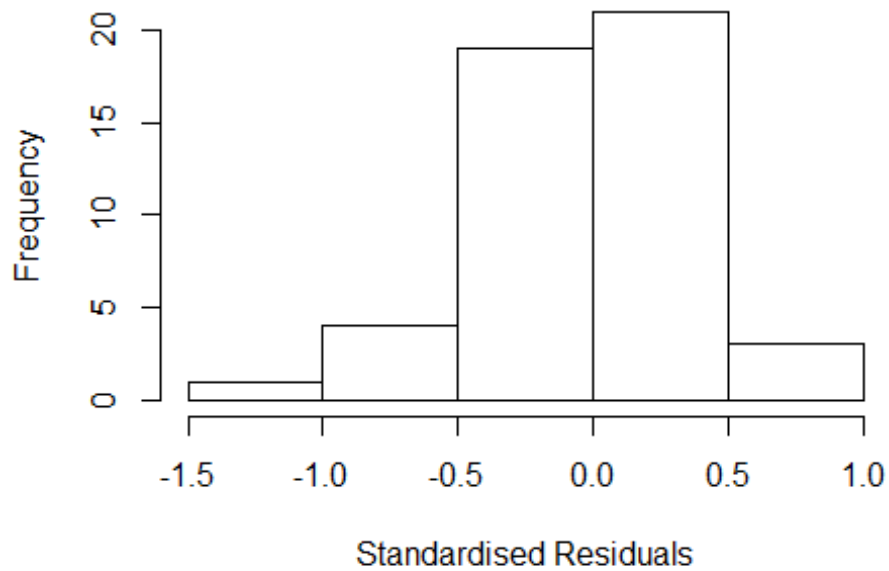
## (Intercept)          9.06e-01    1.19e-08    7.60e+07    < 2e-16 ***
## FirstAuthorFemale1 -8.11e-03    1.39e-01   -6.00e-02    0.95381
## Year2001            2.58e-01    1.81e-08    1.43e+07    < 2e-16 ***
## Year2002           -3.46e-01    2.58e-01   -1.34e+00    0.18773
## Year2003            5.65e-01    4.37e-01    1.29e+00    0.20490
## Year2004            9.01e-02    1.47e-01    6.10e-01    0.54515
## Year2005           -2.66e-02    2.63e-01   -1.00e-01    0.92010
## Year2006           -7.63e-01    1.03e-01   -7.40e+00    1.4e-08 ***
## Year2007            2.32e-01    2.91e-01    8.00e-01    0.43025
## Year2008            3.50e-01    2.12e-01    1.65e+00    0.10736
## Year2009           -6.01e-02    2.41e-01   -2.50e-01    0.80459
## Year2010            2.73e-01    1.26e-01    2.16e+00    0.03751 *
## Year2011            4.38e-01    1.17e-01    3.75e+00    0.00066 ***
## Year2012            5.44e-01    1.81e-01    3.00e+00    0.00500 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.398, Adjusted R-squared:  0.167
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 43 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.406  0.915   0.953   0.924   0.988   0.996
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.08e-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 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

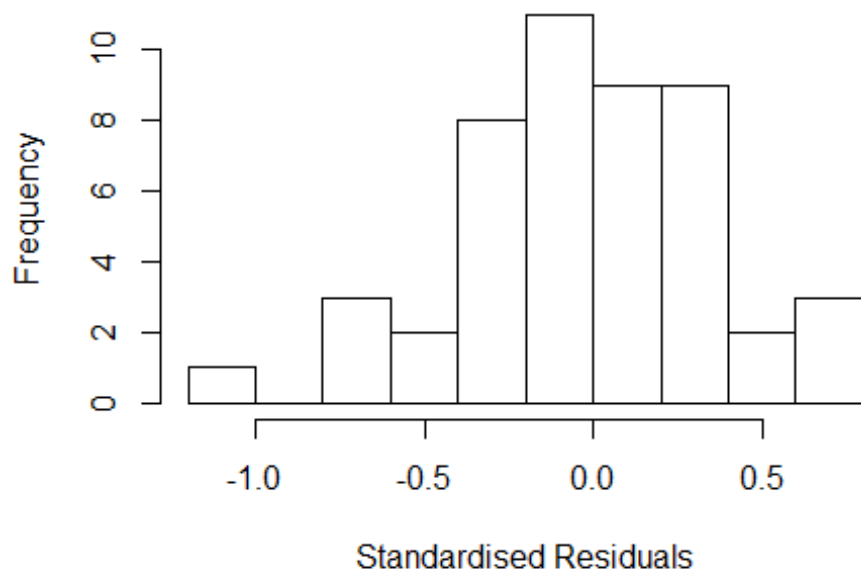
```

Residuals from first author



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale  NaN  1          NaN
## Year              NaN 12          NaN
```

Residuals from last author



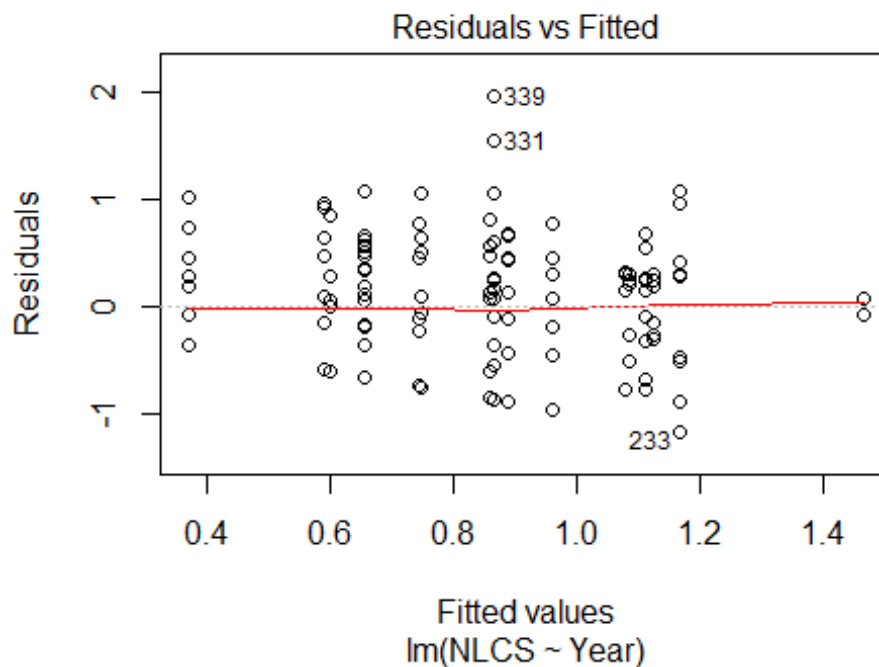
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23920 -0.24508 0.00654 0.23521 0.64997
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9060 0.0000 Inf < 2e-16 ***
## LastAuthorFemale1 0.1405 0.1870 0.75 0.45765
## Year2001 0.2580 0.0000 Inf < 2e-16 ***
## Year2002 -0.4513 0.2628 -1.72 0.09508 .
## Year2003 0.5645 0.4528 1.25 0.22107
## Year2004 -0.0099 0.1447 -0.07 0.94589
## Year2005 -0.0556 0.3039 -0.18 0.85588
## Year2006 -0.7630 0.1033 -7.38 1.5e-08 ***
## Year2007 0.0926 0.3385 0.27 0.78604
## Year2008 0.3332 0.2421 1.38 0.17780
## Year2009 -0.0805 0.2086 -0.39 0.70206
## Year2010 0.2223 0.1069 2.08 0.04510 *
## Year2011 0.3976 0.1063 3.74 0.00068 ***
## Year2012 0.5109 0.1847 2.77 0.00910 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared: 0.418, Adjusted R-squared: 0.195
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 43 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.362 0.909 0.952 0.918 0.988 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 2.08e-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: 48"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3606"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1999 2005
##    2    4
##
## 1999 2005
##    2    4
##
## 1999 2005
##    2    4
## [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: 6"
## [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
##   26   37   14   11   55    8   18    8   10   10    9   13   13   18   10
## 2011 2012
##   29   23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    7    4    5   11    0    9    6    6    6    8    9    9   13    8
## 2011 2012

```

```
## 22 18
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 6 4 4 11 0 8 5 4 4 6 8 7 13 8
## 2011 2012
## 20 18
## [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 = 15, p-value = 0.4
```



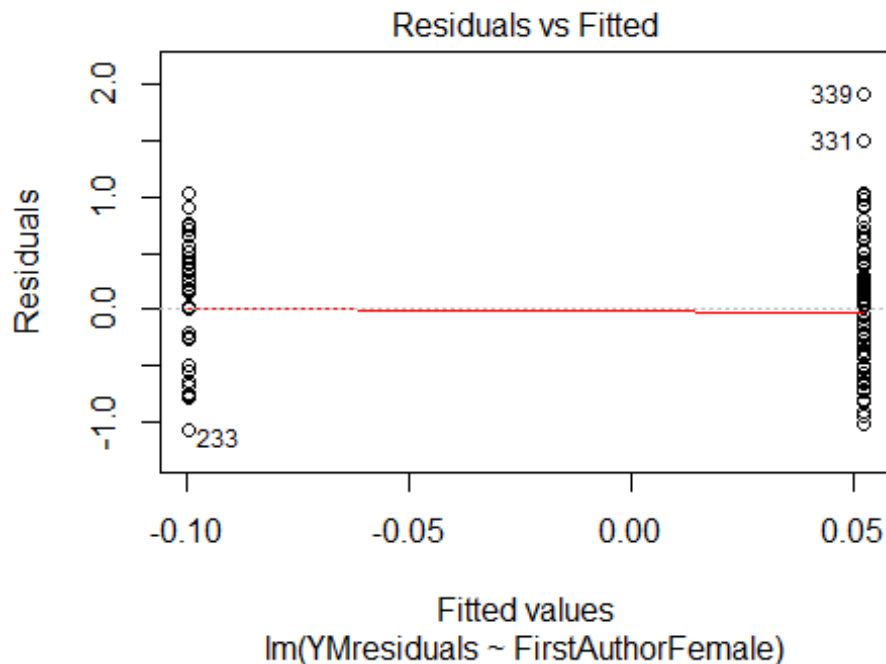
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.18, df = 1, p-value = 0.7

## [1] "Female first author team size 2018 geometric mean: 4.48140474655716"
## [1] "Male first author team size 2018 geometric mean: 4.36449454388689"

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

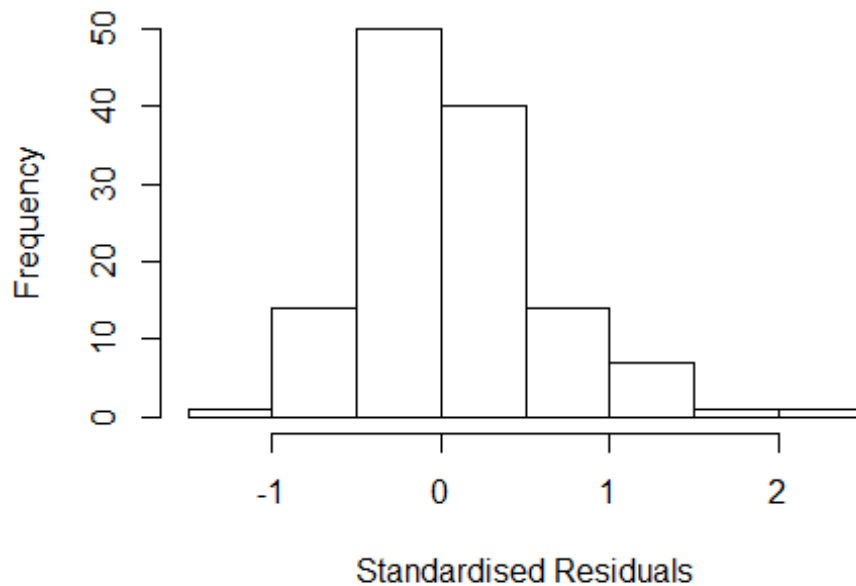
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7.5, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4"
## [1] "Male last author team size 2018 geometric mean: 4.45628831179955"

## 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 = 2.5, 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  4.502  1      2.122
## LastAuthorFemale  3.382  1      1.839
## UniqueAuthors    296.536  4      2.037
## Year              863.479 15      1.253
```

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
## -1.2388 -0.2428 -0.0157 0.2944 2.0263
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.59086 0.28862 2.05 0.04312 *
## FirstAuthorFemale1 0.05123 0.11899 0.43 0.66770
## LastAuthorFemale1 -0.11367 0.11227 -1.01 0.31364
## UniqueAuthors2 0.65845 0.17430 3.78 0.00026 ***
## UniqueAuthors3 0.62644 0.18936 3.31 0.00128 **
## UniqueAuthors4 0.54429 0.37944 1.43 0.15438
## UniqueAuthors5 0.82192 0.20656 3.98 0.00013 ***
## Year1997 -0.15857 0.48123 -0.33 0.74242
## Year1998 0.00333 0.31868 0.01 0.99168
## Year1999 0.16661 0.31491 0.53 0.59787
```

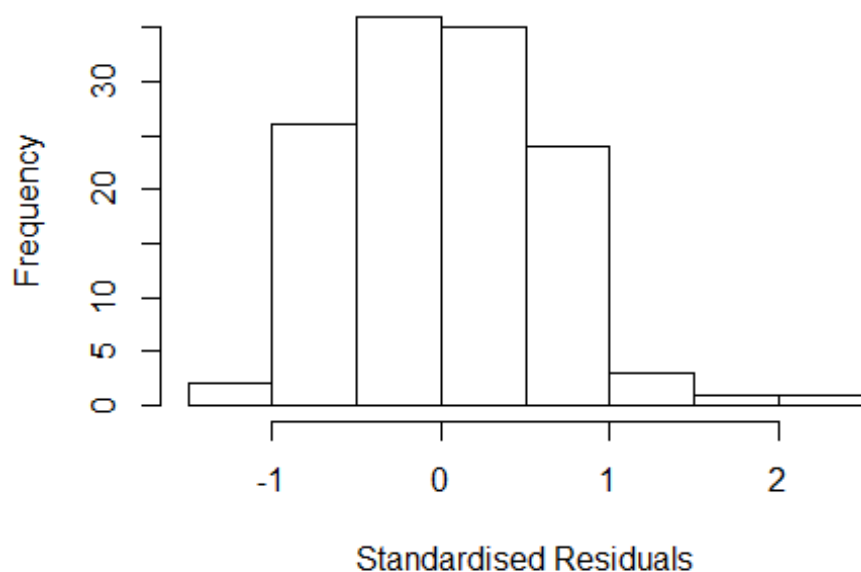


```

## Year2000      -0.43195      0.33899      -1.27      0.20537
## Year2002      0.09178      0.22326      0.41      0.68184
## Year2003     -0.13968      0.29208     -0.48      0.63348
## Year2004     -0.19062      0.42782     -0.45      0.65682
## Year2005     -0.40028      0.28445     -1.41      0.16230
## Year2006     -0.31062      0.42389     -0.73      0.46532
## Year2007     -0.14692      0.35709     -0.41      0.68157
## Year2008     -0.26075      0.52949     -0.49      0.62341
## Year2009     -0.70408      0.24454     -2.88      0.00482 **
## Year2010     -0.17400      0.35062     -0.50      0.62073
## Year2011     -0.34137      0.16413     -2.08      0.03995 *
## Year2012     -0.33191      0.16610     -2.00      0.04825 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.433, Adjusted R-squared:  0.321
## Convergence in 41 IRWLS iterations
##
## Robustness weights:
## observation 128 is an outlier with |weight| = 0 ( < 0.00078);
## 8 weights are ~= 1. The remaining 119 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0551 0.8240 0.9620 0.8690 0.9820 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      7.81e-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 2.810 1      1.676
## LastAuthorFemale 2.157 1      1.469
## Year      5.323 15      1.057

```

Residuals from first and last author



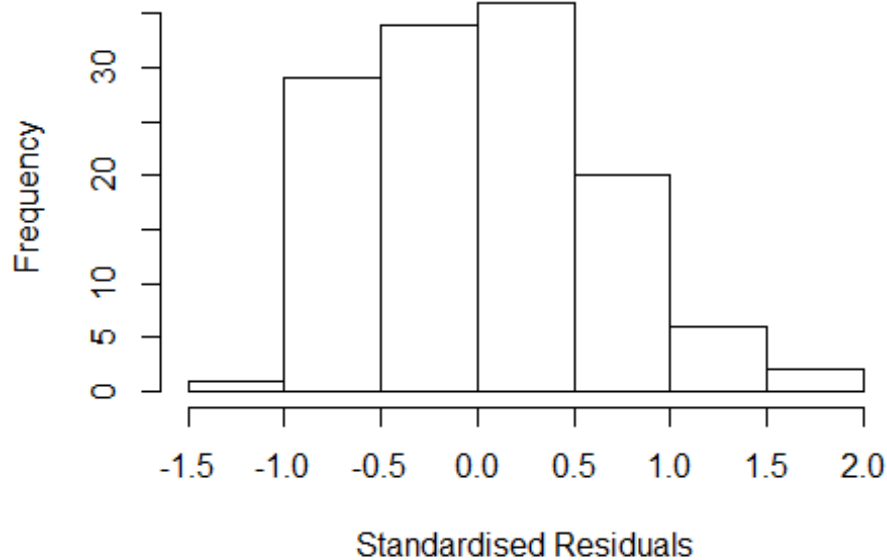
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.09612 -0.45869 0.00759 0.46649 2.17891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.5352 0.1494 10.28 < 2e-16 ***
## FirstAuthorFemale1 -0.0712 0.1388 -0.51 0.6089
## LastAuthorFemale1 -0.2315 0.1416 -1.64 0.1049
## Year1997 -0.6044 0.2730 -2.21 0.0289 *
## Year1998 -0.1985 0.3590 -0.55 0.5814
## Year1999 -0.4482 0.2661 -1.68 0.0949 .
## Year2000 -0.8405 0.2652 -3.17 0.0020 **
## Year2002 -0.4391 0.2680 -1.64 0.1042
## Year2003 -0.3804 0.1468 -2.59 0.0108 *
## Year2004 -1.0555 0.3548 -2.97 0.0036 **
## Year2005 -0.5767 0.3108 -1.86 0.0662 .
## Year2006 -0.7441 0.2645 -2.81 0.0058 **
```

```

## Year2007          -0.3445      0.2277   -1.51    0.1330
## Year2008          -0.3723      0.4283   -0.87    0.3866
## Year2009          -1.0863      0.1697   -6.40    3.9e-09 ***
## Year2010          -0.7179      0.2975   -2.41    0.0175 *
## Year2011          -0.7749      0.1615   -4.80    5.1e-06 ***
## Year2012          -0.6526      0.1950   -3.35    0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.663
## Multiple R-squared:  0.168, Adjusted R-squared:  0.0396
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.258  0.904   0.953   0.927   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.81e-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 2.594 1          1.611
## Year              2.594 15          1.032

```

Residuals from first author



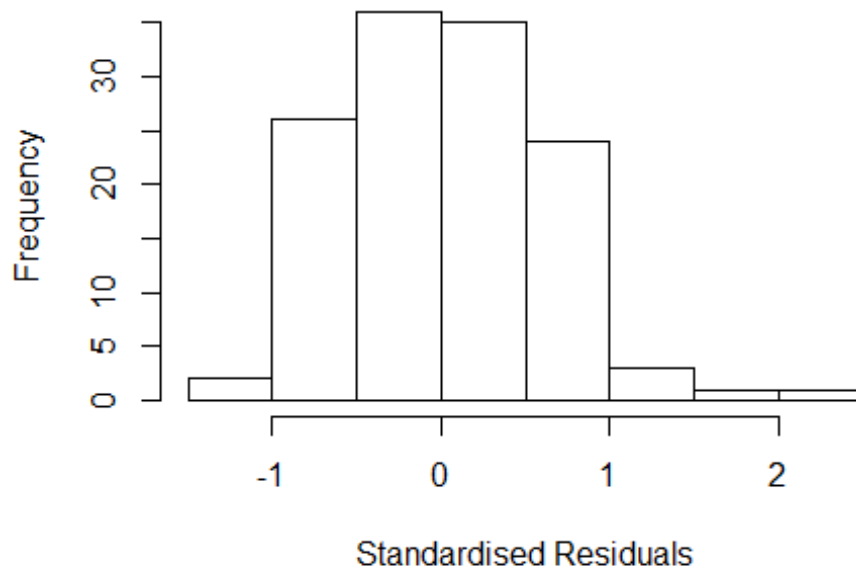
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.03757 -0.48267 0.00957 0.43343 1.98224
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.618 0.143 11.35 < 2e-16 ***
## FirstAuthorFemale1 -0.154 0.131 -1.17 0.24310
## Year1997 -0.671 0.278 -2.41 0.01753 *
## Year1998 -0.435 0.261 -1.67 0.09847 .
## Year1999 -0.587 0.230 -2.55 0.01198 *
## Year2000 -0.997 0.247 -4.04 9.8e-05 ***
## Year2002 -0.581 0.247 -2.35 0.02031 *
## Year2003 -0.462 0.151 -3.06 0.00279 **
## Year2004 -1.135 0.354 -3.21 0.00173 **
## Year2005 -0.773 0.351 -2.20 0.02969 *
## Year2006 -0.869 0.247 -3.51 0.00064 ***
## Year2007 -0.496 0.212 -2.33 0.02137 *
```

```

## Year2008          -0.544      0.404   -1.35   0.18051
## Year2009          -1.217      0.152   -8.01   1.3e-12 ***
## Year2010          -0.841      0.294   -2.86   0.00505 **
## Year2011          -0.888      0.164   -5.40   3.8e-07 ***
## Year2012          -0.770      0.220   -3.50   0.00066 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.688
## Multiple R-squared:  0.142, Adjusted R-squared:  0.0179
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.387  0.909   0.956   0.931   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      7.81e-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 1.844 1          1.358
## Year            1.844 15          1.021

```

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
## -1.0936 -0.4761 0.0105 0.4506 2.2403
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4640 0.0553 26.48 < 2e-16 ***
## LastAuthorFemale1 -0.2557 0.1320 -1.94 0.0553 .
## Year1997 -0.5477 0.2603 -2.10 0.0376 *
## Year1998 -0.1476 0.3412 -0.43 0.6661
## Year1999 -0.3714 0.2309 -1.61 0.1106
## Year2000 -0.7817 0.2334 -3.35 0.0011 **
## Year2002 -0.3704 0.2240 -1.65 0.1011
## Year2003 -0.3467 0.1240 -2.80 0.0061 **
## Year2004 -0.9843 0.3318 -2.97 0.0037 **
## Year2005 -0.4937 0.2596 -1.90 0.0599 .
## Year2006 -0.6903 0.2662 -2.59 0.0108 *
## Year2007 -0.2823 0.1953 -1.45 0.1511
```

```

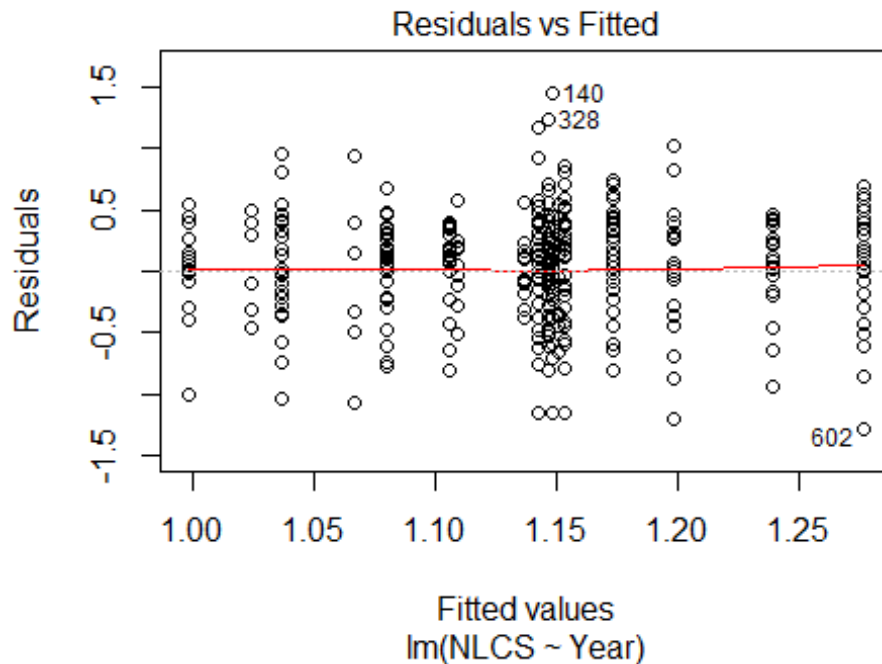
## Year2008          -0.3098      0.4030   -0.77    0.4436
## Year2009          -1.0304      0.1418   -7.27    5.5e-11 ***
## Year2010          -0.6595      0.2781   -2.37    0.0194 *
## Year2011          -0.7334      0.1432   -5.12    1.3e-06 ***
## Year2012          -0.6186      0.2049   -3.02    0.0031 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.167, Adjusted R-squared:  0.0474
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 123 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.229  0.906  0.953  0.927  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      7.81e-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: 128"
## [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]"
##
## 1999 2005
##    2    4
##
## 1999 2005
##    2    4
##
## 1999 2005
##    2    4
## [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: 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: 6"
## [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
##      8   14   14   18   14   24   22   25   28   21   30   37   46   53   55
## 2011 2012
##    71   54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      7    9   13   16   11   14   21   23   24   18   28   34   43   47   50
## 2011 2012
##    66   45
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      6    9   12   13   10   14   19   21   21   15   24   33   39   44   45
## 2011 2012
##    62   43
## [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

```

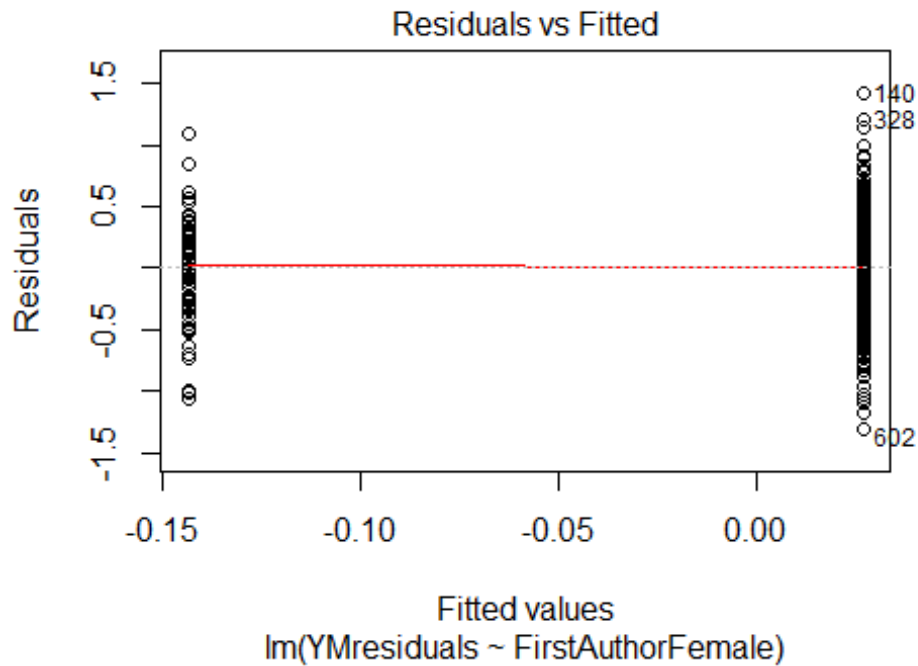
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0074, df = 1, p-value = 0.9

## [1] "Female first author team size 2018 geometric mean: 3.69243372791407"
## [1] "Male first author team size 2018 geometric mean: 3.15107168419581"

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

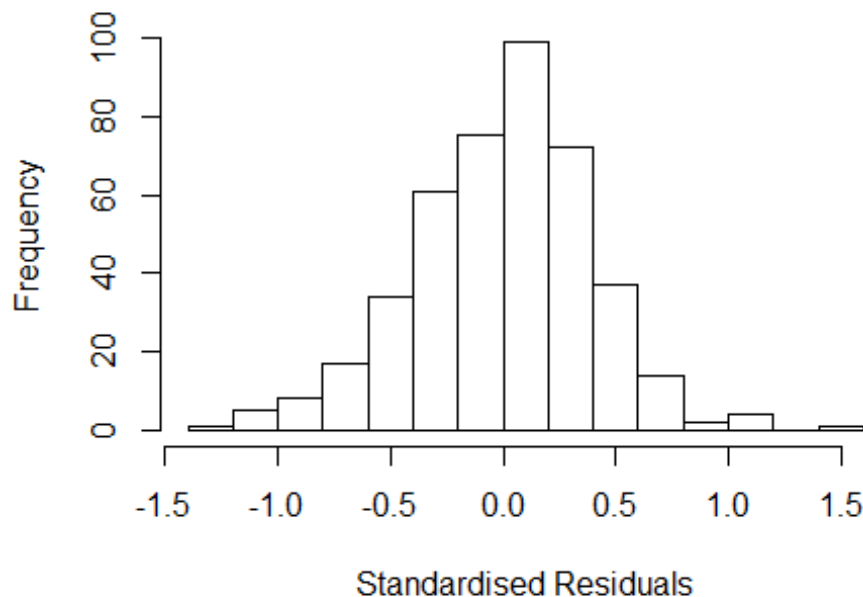
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.78215662420228"
## [1] "Male last author team size 2018 geometric mean: 2.92651831932645"

## 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 = 270, 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 1.389  1          1.179
## LastAuthorFemale  1.204  1          1.097
## UniqueAuthors    3.248  4          1.159
## Year              3.567 16          1.041
```

Residuals from first and last author and team size



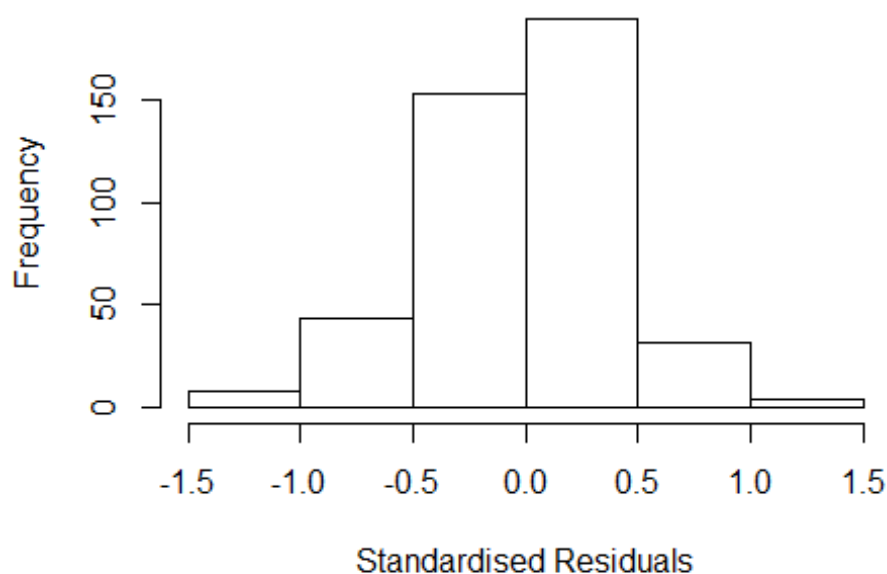
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2310 -0.2628 0.0322 0.2443 1.4241
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99888 0.35575 2.81 0.00523 **
## FirstAuthorFemale1 0.19118 0.05585 3.42 0.00068 ***
## LastAuthorFemale1 -0.00415 0.04762 -0.09 0.93062
## UniqueAuthors2 0.19620 0.07054 2.78 0.00566 **
## UniqueAuthors3 0.16787 0.07467 2.25 0.02510 *
## UniqueAuthors4 0.17333 0.08539 2.03 0.04302 *
## UniqueAuthors5 0.18629 0.08418 2.21 0.02746 *
## Year1997 -0.21160 0.36994 -0.57 0.56764
## Year1998 -0.21563 0.36634 -0.59 0.55646
## Year1999 -0.15116 0.37041 -0.41 0.68342
```

```

## Year2000      -0.15064      0.37041      -0.41      0.68446
## Year2001      -0.18236      0.35410      -0.51      0.60684
## Year2002      -0.02951      0.35925      -0.08      0.93458
## Year2003      -0.20028      0.36020      -0.56      0.57850
## Year2004      -0.09448      0.36238      -0.26      0.79443
## Year2005      -0.14858      0.35432      -0.42      0.67520
## Year2006      -0.23757      0.35879      -0.66      0.50825
## Year2007      -0.26433      0.35299      -0.75      0.45440
## Year2008      -0.14739      0.35112      -0.42      0.67488
## Year2009      -0.14900      0.35222      -0.42      0.67249
## Year2010      -0.11984      0.35340      -0.34      0.73470
## Year2011      -0.14627      0.35004      -0.42      0.67627
## Year2012      -0.00702      0.35034      -0.02      0.98402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.374
## Multiple R-squared:  0.0942, Adjusted R-squared:  0.0453
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 393 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.114  0.871  0.952  0.899  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          2.33e-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 1.225 1      1.107
## LastAuthorFemale  1.093 1      1.045
## Year              1.294 16      1.008

```

Residuals from first and last author



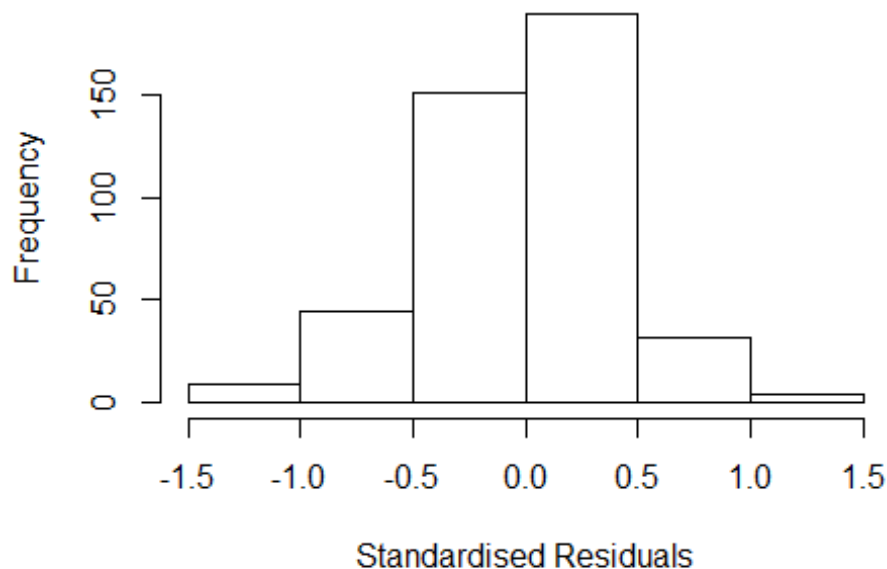
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3688 -0.2436  0.0286  0.2458  1.4483
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1816    0.3060   3.86 0.00013 ***
## FirstAuthorFemale1  0.2011    0.0557   3.61 0.00034 ***
## LastAuthorFemale1 -0.0138    0.0473  -0.29 0.76989
## Year1997          -0.3333    0.3287  -1.01 0.31111
## Year1998          -0.2449    0.3341  -0.73 0.46400
## Year1999          -0.2856    0.3197  -0.89 0.37208
## Year2000          -0.2407    0.3271  -0.74 0.46220
## Year2001          -0.2239    0.3145  -0.71 0.47689
## Year2002          -0.0630    0.3222  -0.20 0.84504
## Year2003          -0.2211    0.3197  -0.69 0.48960
## Year2004          -0.1159    0.3268  -0.35 0.72290
## Year2005          -0.1702    0.3175  -0.54 0.59209
```

```

## Year2006          -0.2720      0.3184   -0.85   0.39341
## Year2007          -0.2774      0.3138   -0.88   0.37719
## Year2008          -0.1789      0.3123   -0.57   0.56701
## Year2009          -0.1691      0.3123   -0.54   0.58854
## Year2010          -0.1441      0.3136   -0.46   0.64598
## Year2011          -0.1632      0.3109   -0.53   0.59984
## Year2012          -0.0272      0.3111   -0.09   0.93029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0672, Adjusted R-squared:  0.0264
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 400 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.106  0.866  0.957  0.899  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      2.33e-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 1.211 1      1.101
## Year              1.211 16      1.006

```

Residuals from first author



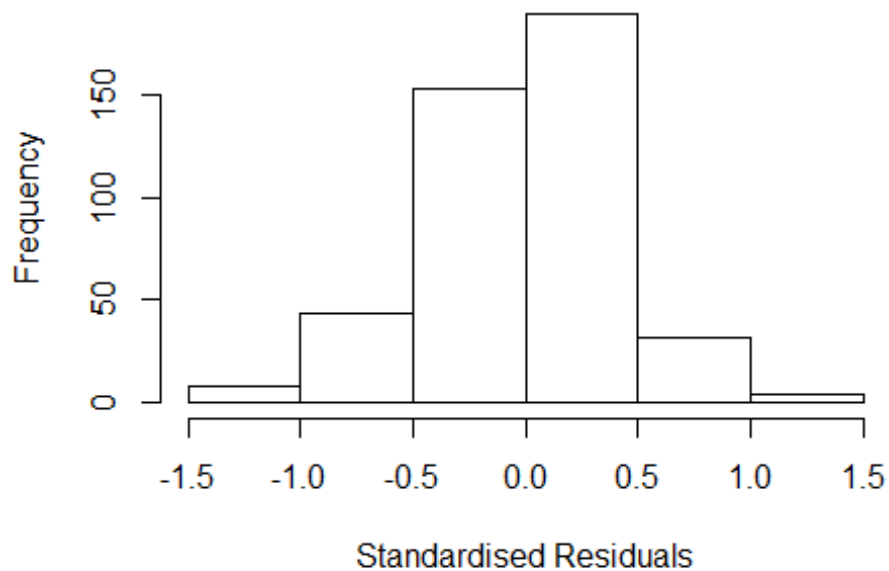
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3880 -0.2447 0.0247 0.2444 1.4443
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1913 0.2941 4.05 6.1e-05 ***
## FirstAuthorFemale1 0.1967 0.0558 3.53 0.00047 ***
## Year1997 -0.3513 0.3268 -1.07 0.28305
## Year1998 -0.2639 0.3277 -0.81 0.42106
## Year1999 -0.2997 0.3111 -0.96 0.33591
## Year2000 -0.2579 0.3199 -0.81 0.42054
## Year2001 -0.2404 0.3063 -0.78 0.43300
## Year2002 -0.0779 0.3140 -0.25 0.80420
## Year2003 -0.2363 0.3111 -0.76 0.44800
## Year2004 -0.1306 0.3195 -0.41 0.68298
## Year2005 -0.1873 0.3100 -0.60 0.54591
## Year2006 -0.2859 0.3101 -0.92 0.35713
```

```

## Year2007          -0.2951      0.3059   -0.96  0.33526
## Year2008          -0.1976      0.3047   -0.65  0.51691
## Year2009          -0.1868      0.3046   -0.61  0.54006
## Year2010          -0.1606      0.3056   -0.53  0.59944
## Year2011          -0.1803      0.3032   -0.59  0.55228
## Year2012          -0.0418      0.3027   -0.14  0.89021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.364
## Multiple R-squared:  0.0692, Adjusted R-squared:  0.0308
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 400 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0798 0.8550 0.9520 0.8930 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      2.33e-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 1.067 1          1.033
## Year              1.067 16          1.002

```


Residuals from last author



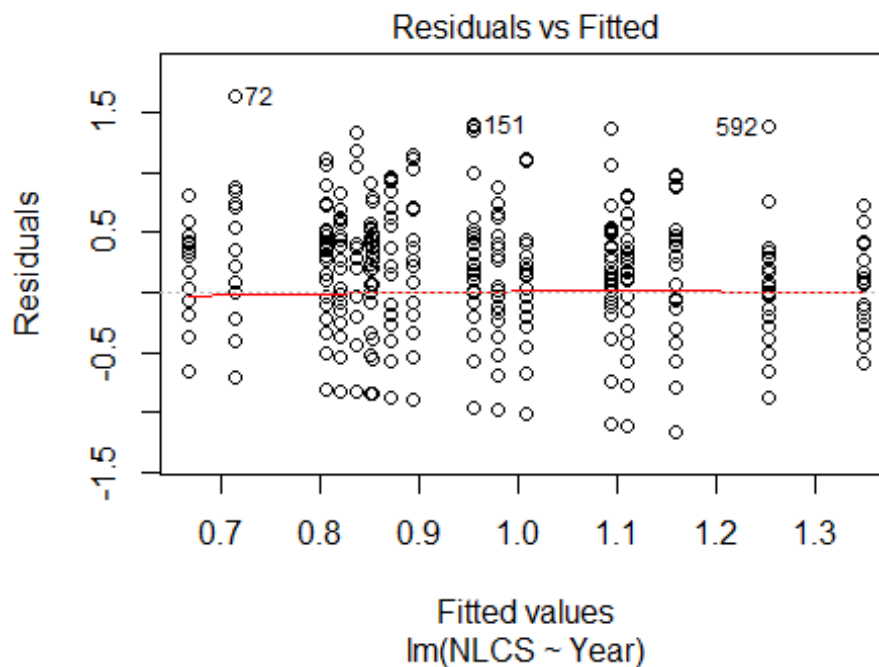
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28559 -0.23829 -0.00232 0.25352 1.44463
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2444 0.3420 3.64 0.00031 ***
## LastAuthorFemale1 0.0412 0.0495 0.83 0.40536
## Year1997 -0.2599 0.3585 -0.72 0.46902
## Year1998 -0.1701 0.3607 -0.47 0.63760
## Year1999 -0.2099 0.3498 -0.60 0.54883
## Year2000 -0.1656 0.3525 -0.47 0.63886
## Year2001 -0.1437 0.3445 -0.42 0.67683
## Year2002 -0.0073 0.3470 -0.02 0.98323
## Year2003 -0.1342 0.3473 -0.39 0.69938
## Year2004 -0.0411 0.3566 -0.12 0.90834
## Year2005 -0.1200 0.3487 -0.34 0.73098
## Year2006 -0.1839 0.3468 -0.53 0.59616
```

```

## Year2007          -0.2282      0.3454   -0.66   0.50909
## Year2008          -0.1257      0.3446   -0.36   0.71544
## Year2009          -0.1210      0.3450   -0.35   0.72599
## Year2010          -0.0699      0.3434   -0.20   0.83871
## Year2011          -0.1102      0.3430   -0.32   0.74813
## Year2012           0.0348      0.3428    0.10   0.91921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.0368, Adjusted R-squared:  -0.00294
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 394 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.109  0.856  0.950  0.895  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      2.33e-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: 430"
## [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
##   24   27   31   33   47   43   26   27   24   28   26   33   33   46   27
## 2011 2012
##   36   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   23   25   27   38   38   23   26   20   23   18   27   27   38   23
## 2011 2012

```

```
## 30 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 19 20 20 24 35 34 22 25 20 23 16 27 24 34 18
## 2011 2012
## 28 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 = 23, df = 16, p-value = 0.1
```



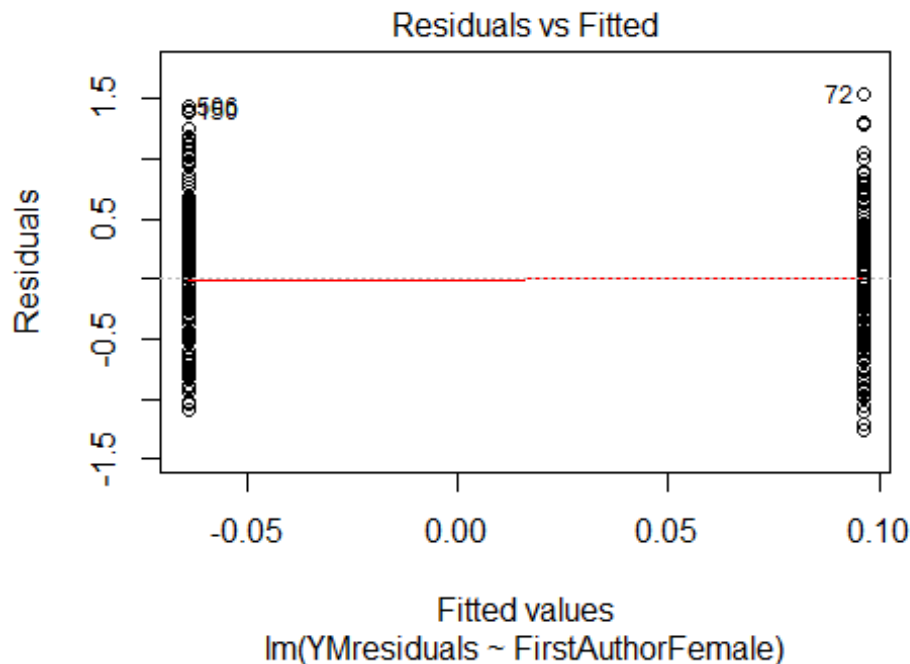
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.06, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 4.4994295404475"
## [1] "Male first author team size 2018 geometric mean: 2.37956557896877"

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

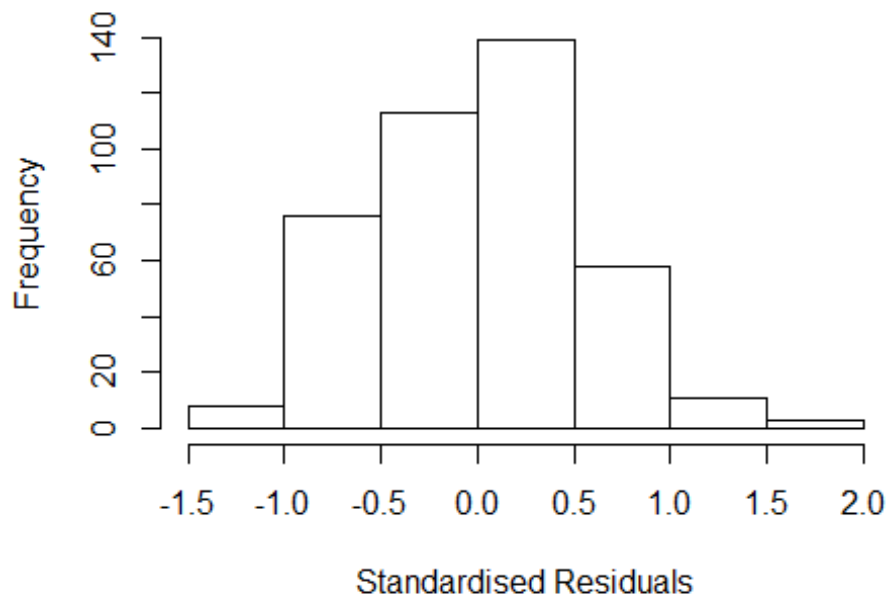
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.95780431090391"
## [1] "Male last author team size 2018 geometric mean: 3.03076363623873"

## 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 = 0.3
## 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 1.551  1      1.245
## LastAuthorFemale  1.390  1      1.179
## UniqueAuthors    3.668  4      1.176
## Year              3.607 16      1.041
```

Residuals from first and last author and team size



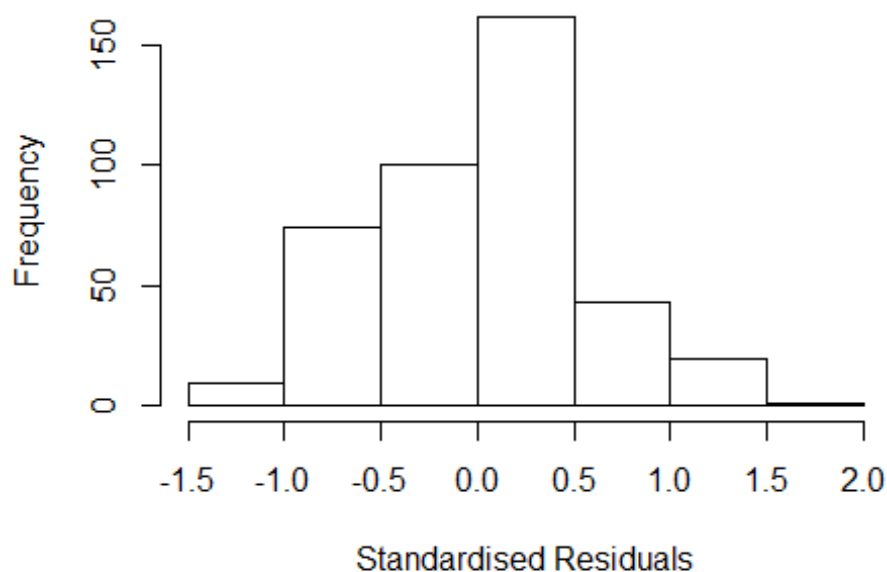
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31961 -0.39905 0.00692 0.34656 1.56506
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.583554 0.165917 3.52 0.00049 ***
## FirstAuthorFemale1 0.067003 0.067845 0.99 0.32398
## LastAuthorFemale1 -0.003334 0.066474 -0.05 0.96003
## UniqueAuthors2 0.163855 0.087122 1.88 0.06076 .
## UniqueAuthors3 0.391481 0.093081 4.21 3.2e-05 ***
## UniqueAuthors4 0.522000 0.107364 4.86 1.7e-06 ***
## UniqueAuthors5 0.591147 0.113179 5.22 2.9e-07 ***
## Year1997 0.030104 0.254977 0.12 0.90608
## Year1998 -0.261094 0.179061 -1.46 0.14562
## Year1999 0.091933 0.178823 0.51 0.60748
```

```

## Year2000      0.144907    0.206945    0.70  0.48421
## Year2001     -0.050586    0.168756   -0.30  0.76452
## Year2002      0.070940    0.186082    0.38  0.70324
## Year2003     -0.000686    0.204432    0.00  0.99733
## Year2004      0.024589    0.198848    0.12  0.90165
## Year2005     -0.152312    0.181622   -0.84  0.40220
## Year2006     -0.059003    0.211532   -0.28  0.78045
## Year2007      0.110160    0.194422    0.57  0.57132
## Year2008      0.203670    0.204967    0.99  0.32101
## Year2009      0.071036    0.176922    0.40  0.68827
## Year2010      0.107667    0.232037    0.46  0.64290
## Year2011      0.280378    0.187500    1.50  0.13564
## Year2012      0.253382    0.191831    1.32  0.18733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.527
## Multiple R-squared:  0.216, Adjusted R-squared:  0.171
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 374 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.357  0.866  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      2.45e-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 1.370 1      1.170
## LastAuthorFemale  1.330 1      1.153
## Year              1.456 16      1.012

```

Residuals from first and last author



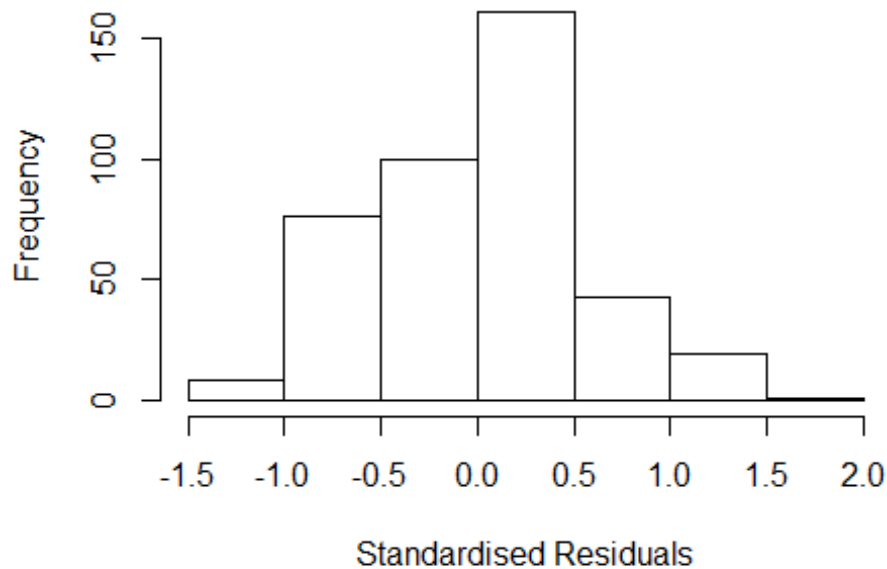
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.267 -0.410 0.058 0.383 1.671
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7229 0.1772 4.08 5.5e-05 ***
## FirstAuthorFemale1 0.1970 0.0673 2.93 0.0036 **
## LastAuthorFemale1 0.0150 0.0675 0.22 0.8236
## Year1997 0.0702 0.2627 0.27 0.7895
## Year1998 -0.2453 0.2087 -1.18 0.2406
## Year1999 0.0397 0.1986 0.20 0.8417
## Year2000 0.1457 0.2137 0.68 0.4957
## Year2001 -0.0370 0.2006 -0.18 0.8538
## Year2002 0.1820 0.2088 0.87 0.3839
## Year2003 0.0255 0.2198 0.12 0.9077
## Year2004 0.0573 0.2224 0.26 0.7968
## Year2005 -0.1112 0.1978 -0.56 0.5741
```

```

## Year2006          0.1023      0.2365      0.43      0.6656
## Year2007          0.1685      0.1996      0.84      0.3991
## Year2008          0.3318      0.2200      1.51      0.1323
## Year2009          0.2077      0.1968      1.06      0.2917
## Year2010          0.2413      0.2900      0.83      0.4059
## Year2011          0.4457      0.1915      2.33      0.0204 *
## Year2012          0.5115      0.2021      2.53      0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.123, Adjusted R-squared:  0.0826
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 380 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.375  0.874   0.952   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      2.45e-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 1.235 1      1.111
## Year              1.235 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
## -1.2585 -0.4084 0.0569 0.3818 1.6654
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7232 0.1777 4.07 5.7e-05 ***
## FirstAuthorFemale1 0.2018 0.0639 3.16 0.0017 **
## Year1997 0.0714 0.2634 0.27 0.7865
## Year1998 -0.2444 0.2085 -1.17 0.2418
## Year1999 0.0428 0.1994 0.21 0.8302
## Year2000 0.1481 0.2139 0.69 0.4892
## Year2001 -0.0357 0.2010 -0.18 0.8592
## Year2002 0.1842 0.2087 0.88 0.3781
## Year2003 0.0293 0.2196 0.13 0.8941
## Year2004 0.0570 0.2225 0.26 0.7978
## Year2005 -0.1069 0.1972 -0.54 0.5880
## Year2006 0.1089 0.2353 0.46 0.6438
```

```

## Year2007          0.1723      0.1984      0.87      0.3858
## Year2008          0.3335      0.2202      1.51      0.1308
## Year2009          0.2112      0.1963      1.08      0.2826
## Year2010          0.2419      0.2908      0.83      0.4060
## Year2011          0.4503      0.1911      2.36      0.0190 *
## Year2012          0.5169      0.2001      2.58      0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.123, Adjusted R-squared:  0.085
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 379 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.376  0.872  0.952  0.911  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      2.45e-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 1.162 1      1.078
## Year      1.162 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
## -1.2148 -0.3987  0.0215  0.3715  1.7916

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7796    0.1679    4.64 4.7e-06 ***
## LastAuthorFemale1 0.0826    0.0640    1.29  0.1980
## Year1997        0.0444    0.2526    0.18  0.8605
## Year1998       -0.2252    0.2109   -1.07  0.2863
## Year1999        0.0287    0.1942    0.15  0.8824
## Year2000        0.1182    0.2077    0.57  0.5697
## Year2001       -0.0429    0.1993   -0.22  0.8298
## Year2002        0.1719    0.2077    0.83  0.4085
## Year2003        0.0109    0.2138    0.05  0.9595
## Year2004        0.0850    0.2187    0.39  0.6978
## Year2005       -0.1468    0.1896   -0.77  0.4393
## Year2006        0.0640    0.2245    0.29  0.7756
## Year2007        0.1973    0.1953    1.01  0.3128
## Year2008        0.3527    0.2139    1.65  0.1001
## Year2009        0.2185    0.1918    1.14  0.2553
## Year2010        0.2921    0.2947    0.99  0.3222
## Year2011        0.4409    0.1844    2.39  0.0173 *
## Year2012        0.5205    0.1928    2.70  0.0072 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.103, Adjusted R-squared:  0.0641
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.312  0.856  0.950  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      2.45e-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: 408"
## [1] ""

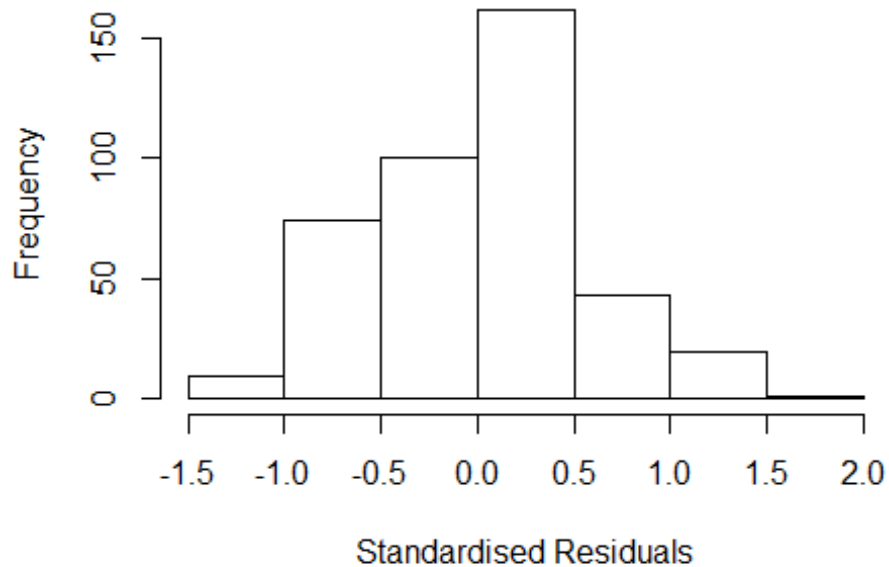
```

```

## [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
##    2    1    4    2    1    8   19   21   22   21   22   47   58   63   42
## 2011 2012
##   46   60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    1    4    1    1    5   17   20   19   14   18   40   50   54   34
## 2011 2012
##   39   48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    0    3    1    1    4   15   20   18   12   16   35   49   51   30
## 2011 2012
##   33   43
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.79091978748794"
## [1] "Male first author team size 2018 geometric mean: 2.7717993116397"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 630, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.58331288953838"
## [1] "Male last author team size 2018 geometric mean: 3.13827764577373"
##
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

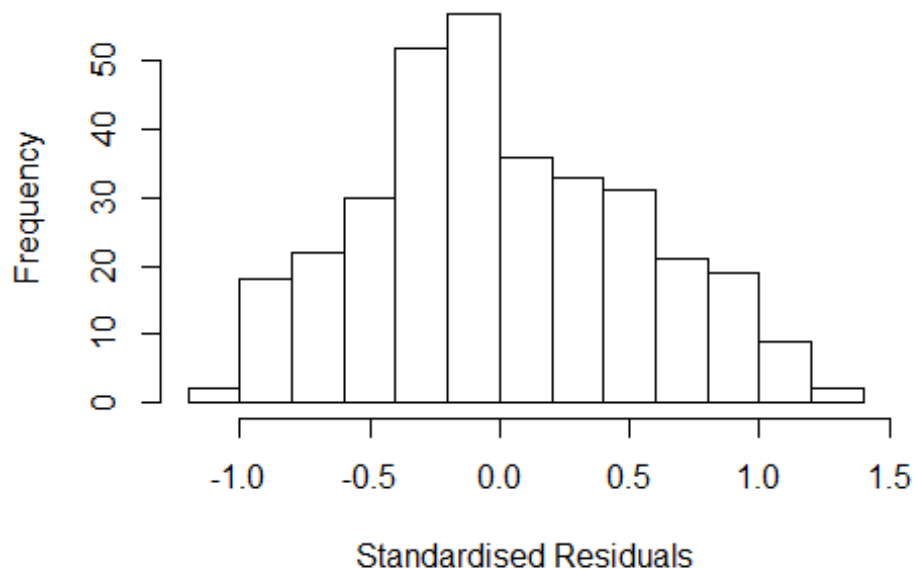
```

Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 610, 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 8.043e+13  1      8.968e+06
## LastAuthorFemale  7.441e+13  1      8.626e+06
## UniqueAuthors    3.742e+28  4      3.729e+03
## Year              7.063e+28 15      9.154e+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
## -1.0925 -0.3280 -0.0613  0.3962  1.3294
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0577    0.1251   8.45 1.1e-15 ***
## FirstAuthorFemale1  0.0871    0.0748   1.16  0.245
## LastAuthorFemale1  0.0789    0.0725   1.09  0.277
## UniqueAuthors2    0.4945    0.0952   5.19 3.8e-07 ***
## UniqueAuthors3    0.5474    0.0948   5.77 1.9e-08 ***
## UniqueAuthors4    0.5316    0.1007   5.28 2.4e-07 ***
## UniqueAuthors5    0.7232    0.1052   6.88 3.4e-11 ***
## Year1998        -1.2621    0.2141  -5.89 9.8e-09 ***
## Year1999         0.2167    0.1461   1.48  0.139
## Year2000         0.1858    0.1204   1.54  0.124
```

```

## Year2001          -0.3610      0.1973    -1.83      0.068 .
## Year2002          -0.8273      0.1814    -4.56     7.3e-06 ***
## Year2003          -0.9452      0.1518    -6.22     1.6e-09 ***
## Year2004          -1.0864      0.1701    -6.39     6.2e-10 ***
## Year2005          -0.7992      0.1730    -4.62     5.6e-06 ***
## Year2006          -1.0548      0.1627    -6.48     3.5e-10 ***
## Year2007          -0.8163      0.1315    -6.21     1.7e-09 ***
## Year2008          -1.0935      0.1171    -9.34    < 2e-16 ***
## Year2009          -0.9353      0.1250    -7.48     7.5e-13 ***
## Year2010          -0.8744      0.1503    -5.82     1.5e-08 ***
## Year2011          -0.9631      0.1335    -7.22     4.1e-12 ***
## Year2012          -0.7756      0.1245    -6.23     1.5e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.258, Adjusted R-squared:  0.207
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 306 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.510  0.855   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          3.01e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.462e+12  1      2.542e+06
## LastAuthorFemale  6.210e+12  1      2.492e+06
## Year              9.403e+12 15      2.707e+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
## -0.9572 -0.4678 -0.0197  0.4522  1.4766
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.72e+00   8.44e-02  2.04e+01 < 2e-16 ***
## FirstAuthorFemale1 1.40e-01   7.44e-02  1.89e+00  0.06 .
## LastAuthorFemale1  8.19e-02   7.29e-02  1.12e+00  0.26
## Year1998        -1.46e+00   1.92e-01 -7.60e+00 3.3e-13 ***
## Year1999         8.13e-02   8.44e-02  9.60e-01  0.34
## Year2000        -4.30e-02   1.09e-07 -3.94e+05 < 2e-16 ***
## Year2001        -5.68e-01   1.43e-01 -3.98e+00 8.7e-05 ***
## Year2002        -1.04e+00   1.50e-01 -6.94e+00 2.3e-11 ***
## Year2003        -1.16e+00   1.16e-01 -1.00e+01 < 2e-16 ***
## Year2004        -1.33e+00   1.54e-01 -8.61e+00 3.5e-16 ***
## Year2005        -1.08e+00   1.53e-01 -7.02e+00 1.4e-11 ***
## Year2006        -1.26e+00   1.36e-01 -9.29e+00 < 2e-16 ***
## Year2007        -9.90e-01   1.16e-01 -8.56e+00 5.0e-16 ***
## Year2008        -1.38e+00   8.73e-02 -1.58e+01 < 2e-16 ***
## Year2009        -1.34e+00   1.09e-01 -1.22e+01 < 2e-16 ***
## Year2010        -1.08e+00   1.31e-01 -8.29e+00 3.3e-15 ***
## Year2011        -1.12e+00   1.37e-01 -8.15e+00 8.5e-15 ***
## Year2012        -1.02e+00   1.04e-01 -9.77e+00 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.626
## Multiple R-squared:  0.118, Adjusted R-squared:  0.0704
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 302 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.557  0.888  0.942  0.921  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      3.01e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01        5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##      trace.lev      mts      compute.rd

```

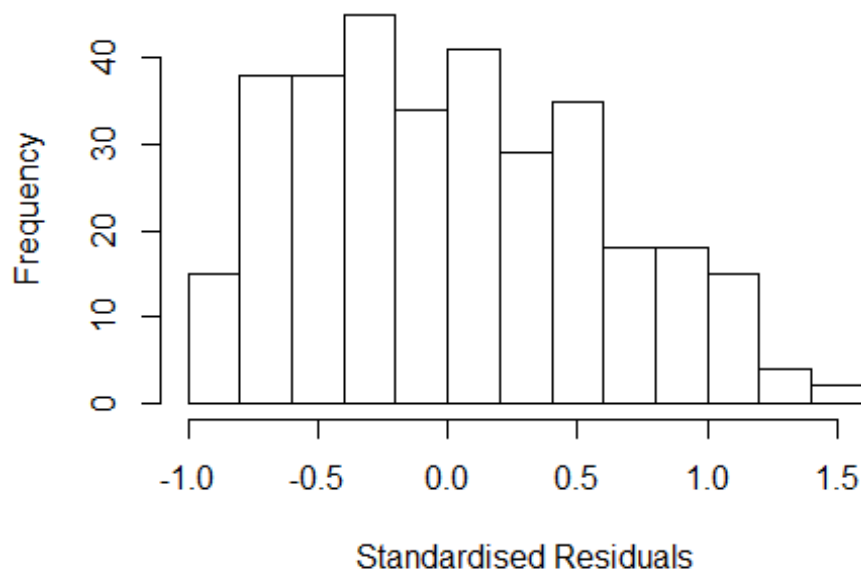


```
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

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

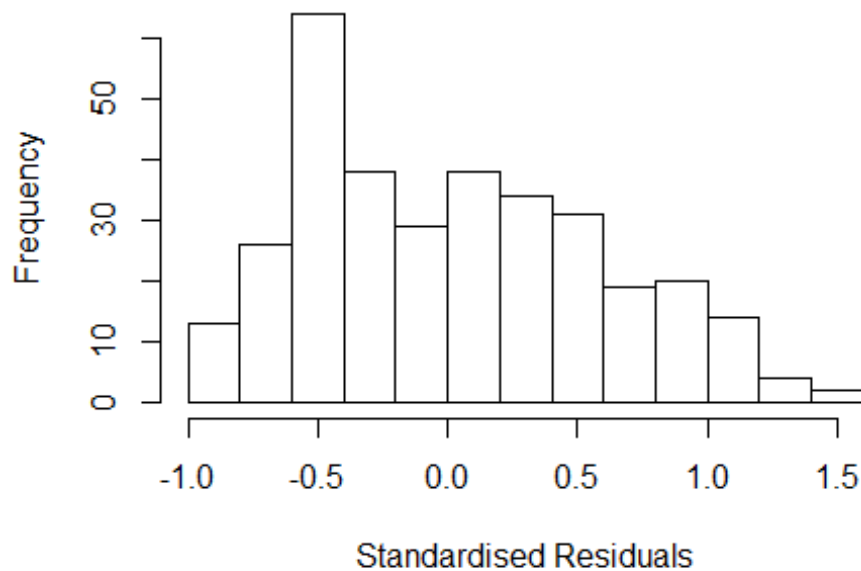
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```

Residuals from first and last author



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1 NaN
## Year NaN 15 NaN
```

Residuals from first author

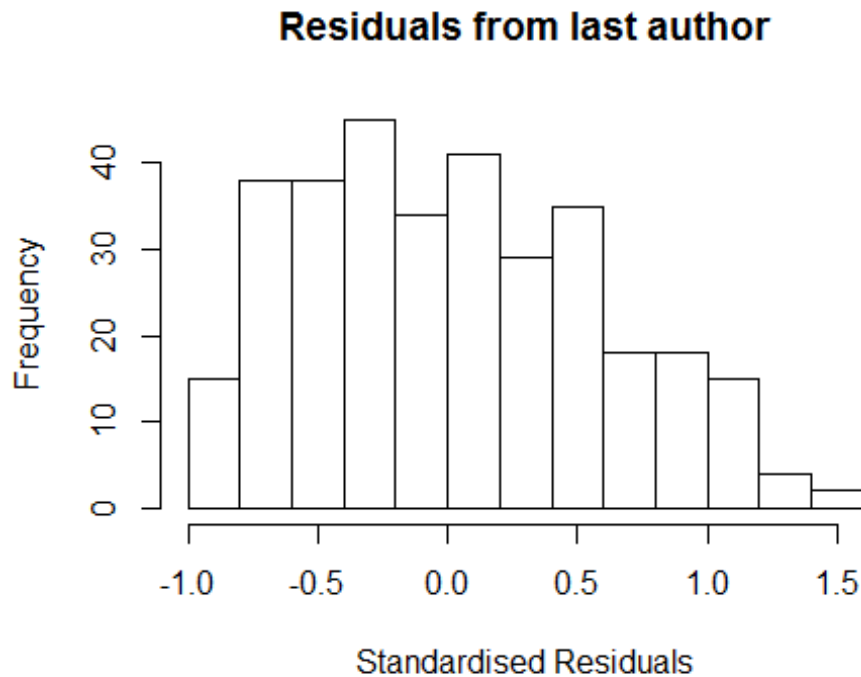


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.93240 -0.45734 -0.00603 0.46490 1.45133
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7820 0.0705 25.28 < 2e-16 ***
## FirstAuthorFemale1 0.1650 0.0705 2.34 0.02 *
## Year1998 -1.5022 0.1960 -7.67 2.2e-13 ***
## Year1999 0.0240 0.0705 0.34 0.73
## Year2000 -0.0430 0.0000 -Inf < 2e-16 ***
## Year2001 -0.6015 0.1385 -4.34 1.9e-05 ***
## Year2002 -1.0671 0.1443 -7.40 1.3e-12 ***
## Year2003 -1.1833 0.1135 -10.42 < 2e-16 ***
## Year2004 -1.3613 0.1467 -9.28 < 2e-16 ***
## Year2005 -1.0903 0.1542 -7.07 9.9e-12 ***
## Year2006 -1.2738 0.1367 -9.32 < 2e-16 ***
## Year2007 -1.0146 0.1159 -8.75 < 2e-16 ***
```

```

## Year2008          -1.4086      0.0815  -17.29  < 2e-16 ***
## Year2009          -1.3700      0.1062  -12.90  < 2e-16 ***
## Year2010          -1.1161      0.1277   -8.74  < 2e-16 ***
## Year2011          -1.1473      0.1322   -8.68  < 2e-16 ***
## Year2012          -1.0463      0.1063   -9.84  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.624
## Multiple R-squared:  0.115, Adjusted R-squared:  0.0703
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 307 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.568  0.894  0.940  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      3.01e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale -1.735e+15  1      NaN
## Year             -1.735e+15 15      NaN

```



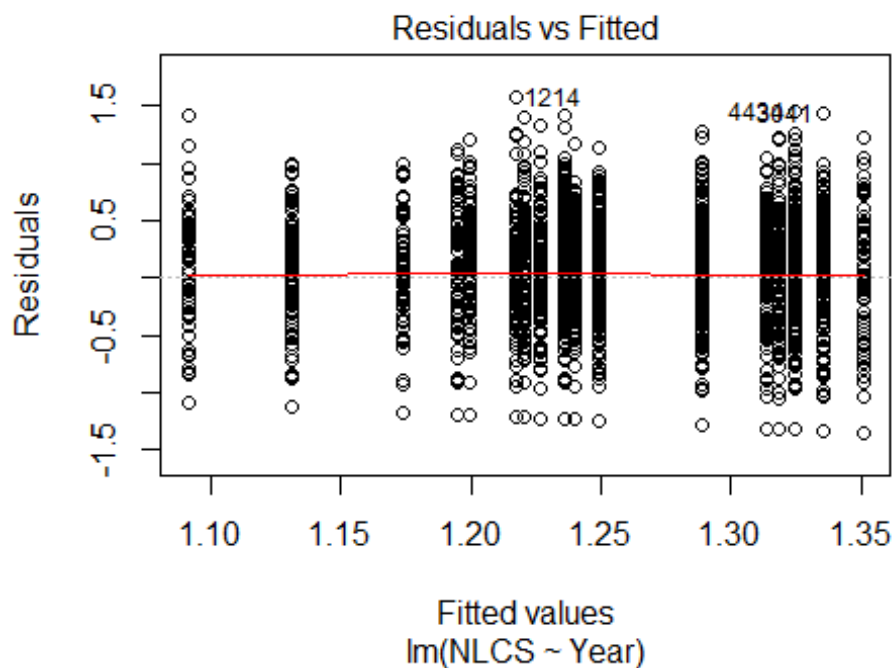
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9164 -0.4403 -0.0206 0.4696 1.4905
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.82e+00 6.92e-02 2.64e+01 < 2e-16 ***
## LastAuthorFemale1 1.22e-01 6.92e-02 1.76e+00 0.07893 .
## Year1998 -1.48e+00 1.92e-01 -7.71e+00 1.7e-13 ***
## Year1999 -1.89e-02 6.92e-02 -2.70e-01 0.78460
## Year2000 -4.30e-02 2.08e-07 -2.06e+05 < 2e-16 ***
## Year2001 -5.82e-01 1.67e-01 -3.49e+00 0.00056 ***
## Year2002 -1.10e+00 1.55e-01 -7.07e+00 9.9e-12 ***
## Year2003 -1.21e+00 1.15e-01 -1.05e+01 < 2e-16 ***
## Year2004 -1.33e+00 1.55e-01 -8.63e+00 3.1e-16 ***
## Year2005 -1.11e+00 1.60e-01 -6.90e+00 2.9e-11 ***
## Year2006 -1.32e+00 1.34e-01 -9.81e+00 < 2e-16 ***
## Year2007 -1.03e+00 1.09e-01 -9.49e+00 < 2e-16 ***
```

```

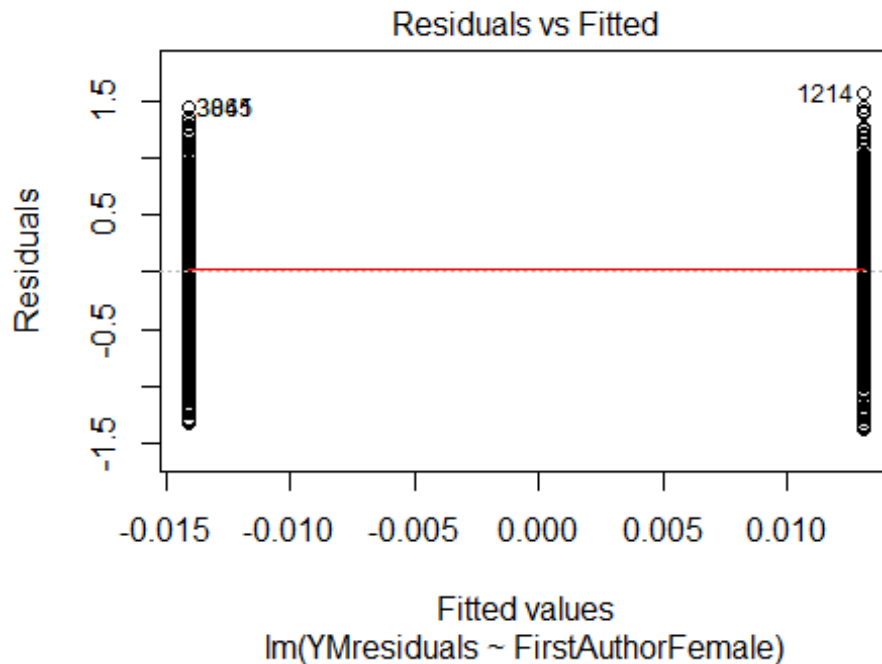
## Year2008          -1.41e+00    8.65e-02 -1.62e+01 < 2e-16 ***
## Year2009          -1.38e+00    1.07e-01 -1.29e+01 < 2e-16 ***
## Year2010          -1.12e+00    1.29e-01 -8.67e+00 2.3e-16 ***
## Year2011          -1.15e+00    1.37e-01 -8.39e+00 1.7e-15 ***
## Year2012          -1.07e+00    1.02e-01 -1.05e+01 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0625
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 310 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.552  0.888  0.947  0.923  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      3.01e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 332"
## [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
## 114 114 141 131 157 182 184 145 195 222 225 247 286 300 359
## 2011 2012
## 403 373
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 84 82 93 104 100 131 90 133 159 158 209 241 223 293
## 2011 2012
## 335 310

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   79   79   75   88   94   89  121   79  122  140  145  192  212  209  268
## 2011 2012
##  316  283
## [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 = 3e-04
```

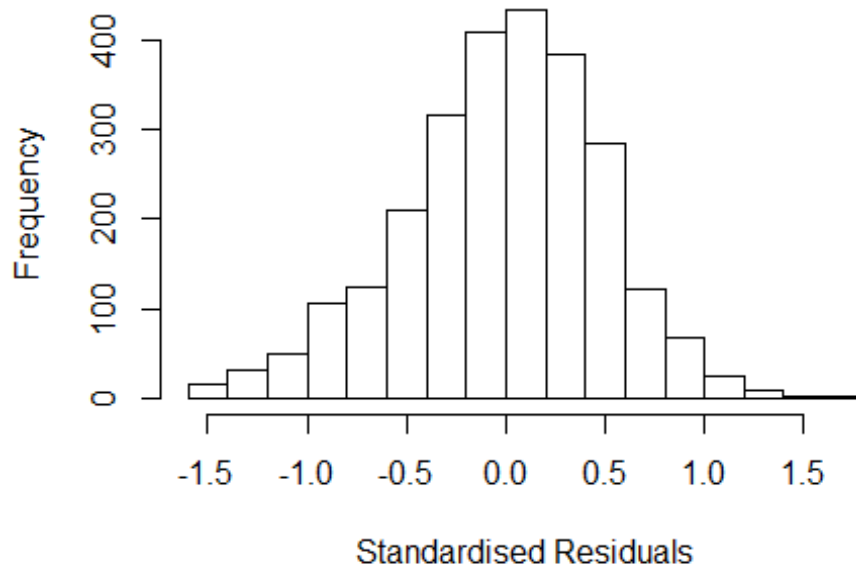


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.07107218934111"
## [1] "Male first author team size 2018 geometric mean: 3.85141353474568"
##
## 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: 3.84579396585799"
## [1] "Male last author team size 2018 geometric mean: 4.0500275881239"
##
## 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 1.133 1      1.065
## LastAuthorFemale  1.127 1      1.061
## UniqueAuthors    1.201 4      1.023
## 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
## -1.5422 -0.3160  0.0166  0.3181  1.7972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.1657    0.0649   17.96 < 2e-16 ***
## FirstAuthorFemale1 -0.0448    0.0209   -2.14  0.03222 *
## LastAuthorFemale1 -0.0111    0.0218   -0.51  0.61136
## UniqueAuthors2     0.2146    0.0433    4.96  7.6e-07 ***
## UniqueAuthors3     0.3500    0.0425    8.24  2.7e-16 ***
## UniqueAuthors4     0.4283    0.0438    9.77 < 2e-16 ***
## UniqueAuthors5     0.5155    0.0435   11.85 < 2e-16 ***
## Year1997         -0.1440    0.0900   -1.60  0.10994
## Year1998         -0.2487    0.0808   -3.08  0.00211 **
## Year1999         -0.3013    0.0827   -3.64  0.00028 ***
```

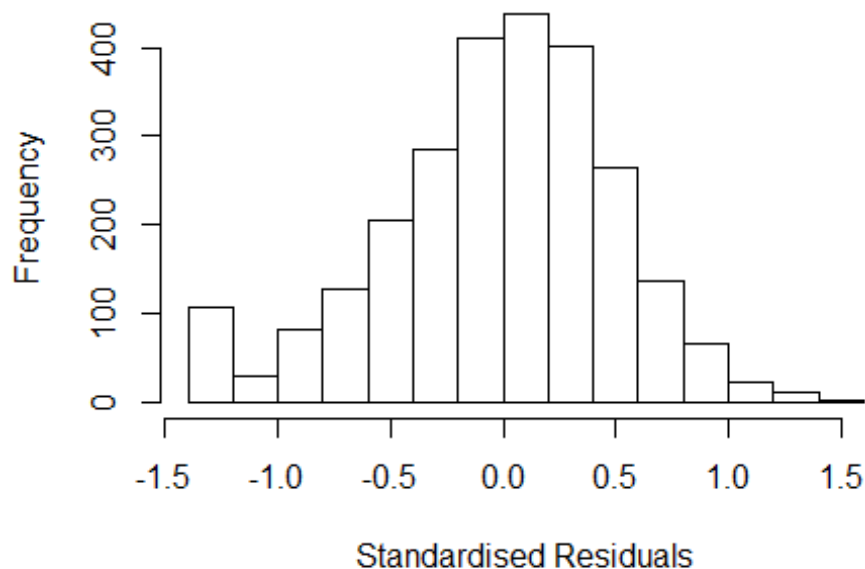


```

## Year2000          -0.2831      0.0773   -3.66  0.00026 ***
## Year2001          -0.1147      0.0918   -1.25  0.21149
## Year2002          -0.1188      0.0753   -1.58  0.11479
## Year2003          -0.2532      0.0800   -3.16  0.00157 **
## Year2004          -0.2319      0.0770   -3.01  0.00264 **
## Year2005          -0.2540      0.0746   -3.41  0.00067 ***
## Year2006          -0.1277      0.0707   -1.81  0.07105 .
## Year2007          -0.2287      0.0665   -3.44  0.00060 ***
## Year2008          -0.2125      0.0692   -3.07  0.00216 **
## Year2009          -0.1390      0.0667   -2.08  0.03719 *
## Year2010          -0.1684      0.0669   -2.52  0.01183 *
## Year2011          -0.2116      0.0684   -3.09  0.00201 **
## Year2012          -0.1503      0.0672   -2.24  0.02530 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.111, Adjusted R-squared:  0.103
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2371 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.867  0.949  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          3.86e-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 1.139 1          1.067
## LastAuthorFemale  1.135 1          1.065
## Year              1.051 16          1.002

```

Residuals from first and last author



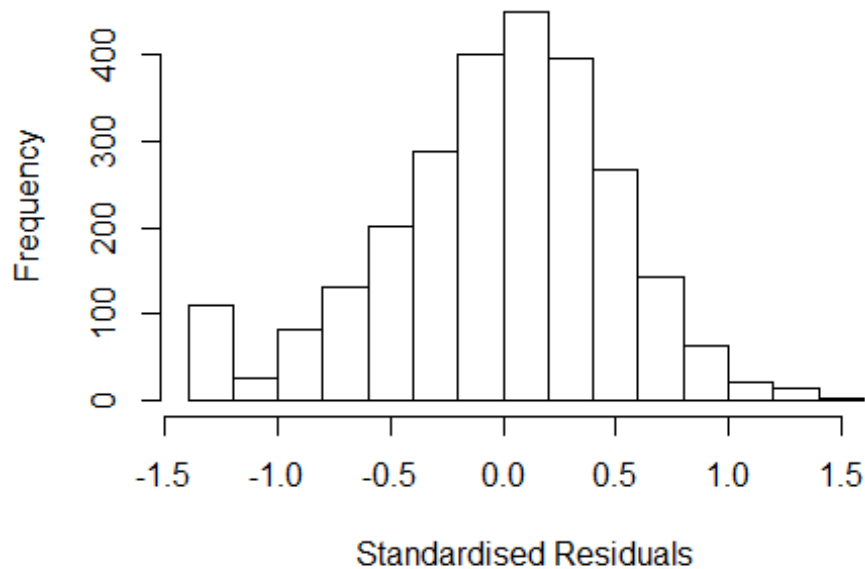
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3776 -0.3361 0.0207 0.3269 1.5565
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36673 0.06396 21.37 <2e-16 ***
## FirstAuthorFemale1 -0.01562 0.02180 -0.72 0.4736
## LastAuthorFemale1 -0.03318 0.02276 -1.46 0.1451
## Year1997 -0.11003 0.09972 -1.10 0.2699
## Year1998 -0.15664 0.08736 -1.79 0.0731 .
## Year1999 -0.23805 0.08926 -2.67 0.0077 **
## Year2000 -0.20852 0.08265 -2.52 0.0117 *
## Year2001 -0.05823 0.09160 -0.64 0.5250
## Year2002 -0.02258 0.07899 -0.29 0.7750
## Year2003 -0.13226 0.08606 -1.54 0.1245
## Year2004 -0.11223 0.08115 -1.38 0.1668
## Year2005 -0.11906 0.07770 -1.53 0.1256
```

```

## Year2006          0.00324    0.07478    0.04    0.9654
## Year2007          -0.08474    0.07055   -1.20    0.2298
## Year2008          -0.08053    0.07211   -1.12    0.2642
## Year2009           0.01080    0.07137    0.15    0.8798
## Year2010          -0.03089    0.07150   -0.43    0.6658
## Year2011          -0.05617    0.07235   -0.78    0.4377
## Year2012           0.01084    0.07038    0.15    0.8777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.477
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.0108
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 233 weights are ~= 1. The remaining 2358 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.862  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      3.86e-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 1.028 1      1.014
## Year              1.028 16      1.001

```

Residuals from first author



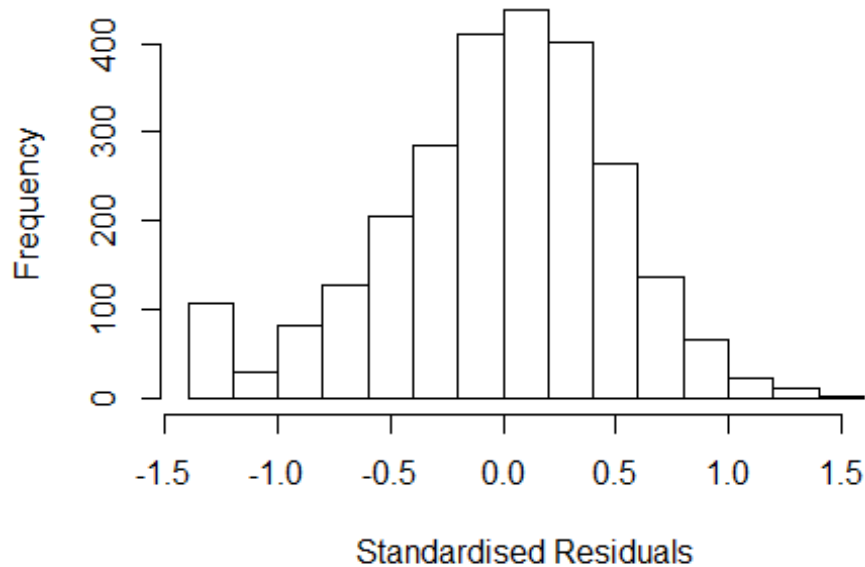
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3714 -0.3350 0.0197 0.3316 1.5615
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.35902 0.06426 21.15 <2e-16 ***
## FirstAuthorFemale1 -0.02709 0.02075 -1.31 0.1918
## Year1997 -0.10657 0.09998 -1.07 0.2866
## Year1998 -0.15632 0.08758 -1.78 0.0744 .
## Year1999 -0.23611 0.08977 -2.63 0.0086 **
## Year2000 -0.20729 0.08311 -2.49 0.0127 *
## Year2001 -0.05876 0.09177 -0.64 0.5221
## Year2002 -0.02108 0.07930 -0.27 0.7904
## Year2003 -0.12949 0.08683 -1.49 0.1360
## Year2004 -0.10956 0.08151 -1.34 0.1790
## Year2005 -0.11526 0.07815 -1.47 0.1404
## Year2006 0.00192 0.07511 0.03 0.9796
```

```

## Year2007          -0.08517    0.07098   -1.20    0.2303
## Year2008          -0.08012    0.07256   -1.10    0.2696
## Year2009           0.01183    0.07173    0.16    0.8691
## Year2010          -0.03202    0.07180   -0.45    0.6557
## Year2011          -0.05470    0.07278   -0.75    0.4523
## Year2012           0.01242    0.07080    0.18    0.8608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.477
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.0102
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 229 weights are ~= 1. The remaining 2362 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.263  0.861  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      3.86e-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 1.025 1          1.012
## 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
## -1.3727 -0.3342 0.0218 0.3264 1.5622
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36284 0.06381 21.36 <2e-16 ***
## LastAuthorFemale1 -0.03897 0.02166 -1.80 0.0722 .
## Year1997 -0.11127 0.09978 -1.12 0.2649
## Year1998 -0.15602 0.08766 -1.78 0.0752 .
## Year1999 -0.23972 0.08921 -2.69 0.0073 **
## Year2000 -0.20945 0.08269 -2.53 0.0114 *
## Year2001 -0.05830 0.09172 -0.64 0.5251
## Year2002 -0.02349 0.07908 -0.30 0.7665
## Year2003 -0.13406 0.08587 -1.56 0.1186
## Year2004 -0.11527 0.08083 -1.43 0.1540
## Year2005 -0.12084 0.07766 -1.56 0.1198
## Year2006 0.00134 0.07476 0.02 0.9857
```

```

## Year2007          -0.08594      0.07052    -1.22    0.2231
## Year2008          -0.08160      0.07205    -1.13    0.2575
## Year2009           0.00986      0.07138     0.14    0.8901
## Year2010          -0.03253      0.07143    -0.46    0.6488
## Year2011          -0.05731      0.07235    -0.79    0.4284
## Year2012           0.00892      0.07028     0.13    0.8990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.477
## Multiple R-squared:  0.0175, Adjusted R-squared:  0.011
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 241 weights are ~= 1. The remaining 2350 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.861  0.946  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      3.86e-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: 2591"
## [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
##    2    4    9    9    5   13    5   11    5    6    8    8    8    7    3
## 2011 2012
##    6    5
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    0    7    7    1   11    4   11    5    6    6    7    8    7    3
## 2011 2012

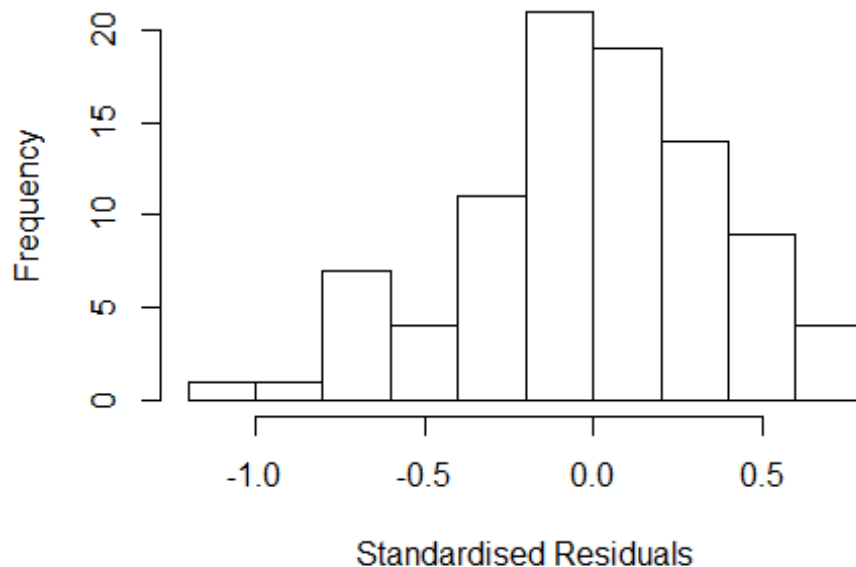
```

```

##      6      2
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      1      0      7      7      1     10      4     11      5      6      6      7      8      7      3
## 2011 2012
##      6      2
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 4"
## [1] "Male first author team size 2018 geometric mean: 2.44948974278318"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 0.7
## 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: 3.46410161513775"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, 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    4.508  1         2.123
## LastAuthorFemale     4.539  1         2.130
## UniqueAuthors       880.366  4         2.334
## Year                 4583.863 15         1.324

```


Residuals from first and last author and team size



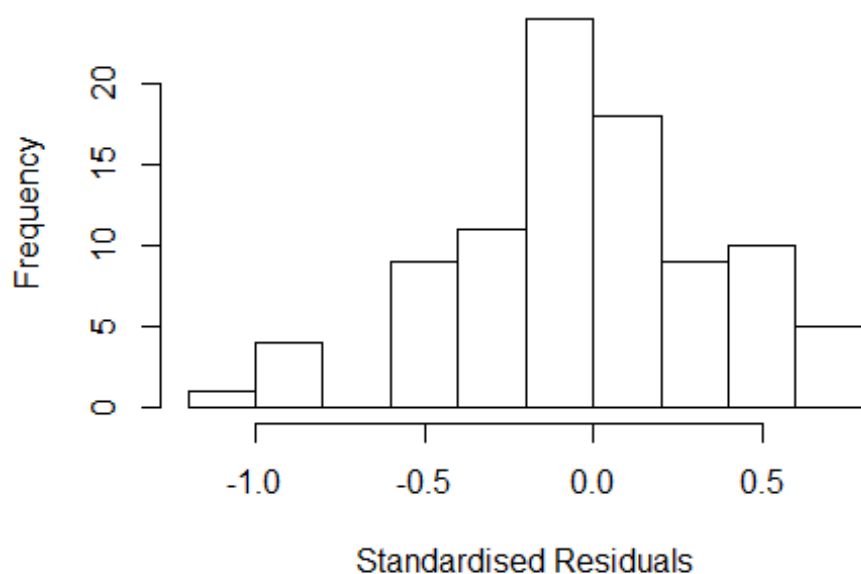
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0968 -0.2187 0.0117 0.2460 0.7448
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01e+00 1.68e-01 6.00e+00 8.2e-08 ***
## FirstAuthorFemale1 6.66e-02 1.18e-01 5.70e-01 0.5729
## LastAuthorFemale1 3.78e-02 1.19e-01 3.20e-01 0.7522
## UniqueAuthors2 2.75e-01 1.50e-01 1.83e+00 0.0719 .
## UniqueAuthors3 2.57e-01 1.68e-01 1.53e+00 0.1306
## UniqueAuthors4 4.90e-01 2.73e-01 1.79e+00 0.0772 .
## UniqueAuthors5 4.25e-01 2.26e-01 1.89e+00 0.0636 .
## Year1998 4.03e-01 2.40e-01 1.68e+00 0.0977 .
## Year1999 1.09e-01 2.50e-01 4.40e-01 0.6638
## Year2000 5.06e-01 5.08e-08 9.96e+06 < 2e-16 ***
```

```

## Year2001      1.98e-02   1.45e-01   1.40e-01   0.8919
## Year2002     -4.41e-01   1.95e-01  -2.26e+00   0.0272 *
## Year2003     -4.50e-02   1.42e-01  -3.20e-01   0.7525
## Year2004     -1.78e-01   2.00e-01  -8.90e-01   0.3760
## Year2005     -5.75e-02   1.54e-01  -3.70e-01   0.7093
## Year2006     -2.29e-01   2.13e-01  -1.07e+00   0.2863
## Year2007     -3.38e-01   2.71e-01  -1.25e+00   0.2166
## Year2008      2.93e-01   3.65e-01   8.00e-01   0.4249
## Year2009     -3.75e-01   1.81e-01  -2.07e+00   0.0421 *
## Year2010      2.37e-01   1.36e-01   1.74e+00   0.0867 .
## Year2011     -8.32e-01   1.84e-01  -4.53e+00   2.4e-05 ***
## Year2012     -4.14e-01   1.47e-01  -2.81e+00   0.0064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.44,   Adjusted R-squared:  0.27
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 78 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.352  0.852  0.953   0.899  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      1.10e-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 4.221 1          2.054
## LastAuthorFemale  2.840 1          1.685
## Year              11.098 15          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
## -1.1221 -0.2387 -0.0135 0.2153 0.7327
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27e+00 4.07e-08 3.11e+07 < 2e-16 ***
## FirstAuthorFemale1 6.29e-02 1.16e-01 5.40e-01 0.590
## LastAuthorFemale1 3.46e-02 1.04e-01 3.30e-01 0.739
## Year1998 2.71e-01 2.08e-01 1.30e+00 0.197
## Year1999 -2.66e-03 1.96e-01 -1.00e-02 0.989
## Year2000 5.06e-01 5.13e-08 9.86e+06 < 2e-16 ***
## Year2001 -1.57e-02 1.25e-01 -1.30e-01 0.901
## Year2002 -4.63e-01 2.50e-01 -1.85e+00 0.068 .
## Year2003 -4.33e-02 1.26e-01 -3.40e-01 0.732
## Year2004 -1.32e-01 1.81e-01 -7.30e-01 0.470
## Year2005 -3.36e-02 1.02e-01 -3.30e-01 0.743
## Year2006 -1.69e-01 2.12e-01 -8.00e-01 0.427
```

```

## Year2007      -3.75e-01   2.90e-01 -1.29e+00   0.200
## Year2008      3.26e-01   2.53e-01  1.29e+00   0.202
## Year2009     -3.90e-01   1.80e-01 -2.17e+00   0.033 *
## Year2010      6.60e-02   5.35e-02  1.23e+00   0.221
## Year2011     -7.90e-01   1.36e-01 -5.79e+00  1.6e-07 ***
## Year2012     -3.22e-01   9.55e-03 -3.37e+01  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.353, Adjusted R-squared:  0.203
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 78 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.416  0.848  0.947  0.906  0.993  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          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-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 3.757 1      1.938
## Year      3.757 15      1.045

## [1] "List of 0 outliers 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
## -1.1071 -0.2436 -0.0131  0.2138  0.7269

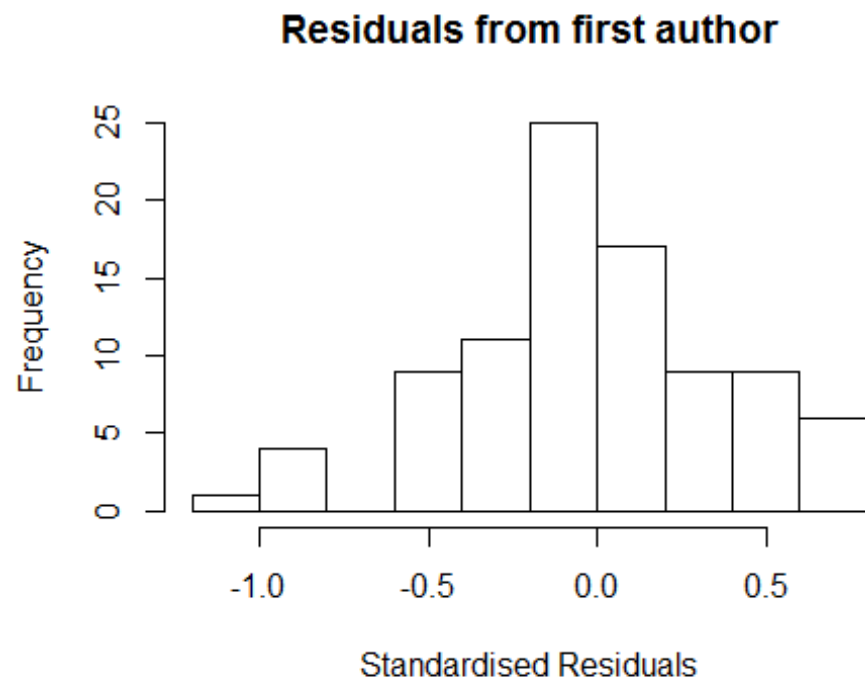
```

```

##
## Coefficients:
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.27e+00   3.73e-08  3.40e+07 < 2e-16 ***
## FirstAuthorFemale1 6.67e-02   1.17e-01   5.70e-01   0.571
## Year1998        2.77e-01   2.06e-01   1.35e+00   0.183
## Year1999       -2.05e-03   1.96e-01  -1.00e-02   0.992
## Year2000        5.06e-01   4.01e-08   1.26e+07 < 2e-16 ***
## Year2001       -9.88e-03   1.21e-01  -8.00e-02   0.935
## Year2002       -4.53e-01   2.57e-01  -1.77e+00   0.081 .
## Year2003       -2.34e-02   1.04e-01  -2.20e-01   0.823
## Year2004       -1.06e-01   1.49e-01  -7.10e-01   0.480
## Year2005       -1.77e-02   9.95e-02  -1.80e-01   0.859
## Year2006       -1.51e-01   2.04e-01  -7.40e-01   0.463
## Year2007       -3.71e-01   2.97e-01  -1.25e+00   0.215
## Year2008        3.41e-01   2.50e-01   1.37e+00   0.176
## Year2009       -3.89e-01   1.84e-01  -2.11e+00   0.038 *
## Year2010        7.75e-02   5.14e-02   1.51e+00   0.136
## Year2011       -7.76e-01   1.42e-01  -5.46e+00  6.1e-07 ***
## Year2012       -3.22e-01   9.55e-03  -3.37e+01 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.354, Adjusted R-squared:  0.214
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 77 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.421  0.846  0.947  0.903  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      1.10e-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"

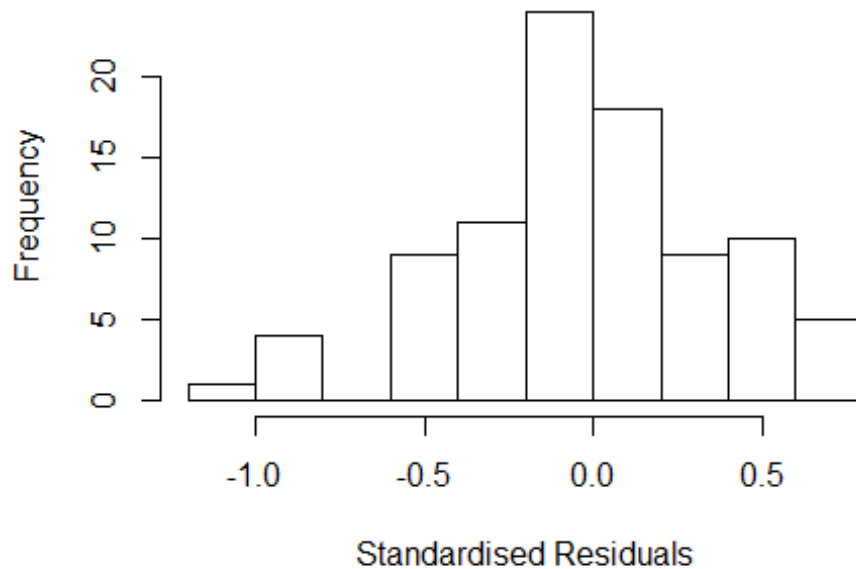
```

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



```
##           GVIF Df GVIF^(1/(2*Df))  
## LastAuthorFemale 2.849 1          1.688  
## Year             2.849 15          1.036
```

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
## -1.08786 -0.23853 -0.00688 0.22794 0.71536
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27e+00 0.00e+00 Inf < 2e-16 ***
## LastAuthorFemale1 4.05e-02 1.04e-01 3.90e-01 0.696
## Year1998 2.81e-01 2.08e-01 1.35e+00 0.180
## Year1999 -8.46e-04 1.98e-01 0.00e+00 0.997
## Year2000 5.06e-01 1.50e-08 3.37e+07 < 2e-16 ***
## Year2001 1.64e-03 1.20e-01 1.00e-02 0.989
## Year2002 -4.61e-01 2.49e-01 -1.85e+00 0.069 .
## Year2003 -2.08e-02 1.16e-01 -1.80e-01 0.858
## Year2004 -1.00e-01 1.63e-01 -6.10e-01 0.542
## Year2005 -3.57e-02 1.02e-01 -3.50e-01 0.728
## Year2006 -1.24e-01 1.60e-01 -7.70e-01 0.443
## Year2007 -3.49e-01 3.09e-01 -1.13e+00 0.262
```

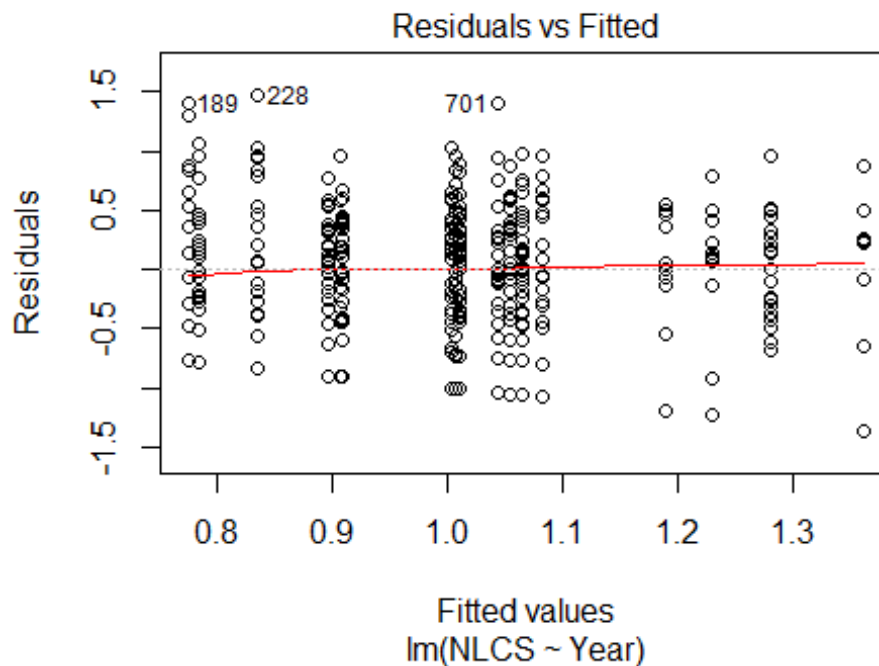
```

## Year2008          3.48e-01  2.53e-01  1.38e+00  0.172
## Year2009         -3.53e-01  1.53e-01 -2.31e+00  0.024 *
## Year2010          6.41e-02  5.24e-02  1.22e+00  0.225
## Year2011         -7.80e-01  1.29e-01 -6.06e+00  5.2e-08 ***
## Year2012         -3.22e-01  9.55e-03 -3.37e+01  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.352, Adjusted R-squared:  0.211
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 79 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.422  0.842  0.949  0.902  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      1.10e-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: 91"
## [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
##   32   37   28   37   51   53   50   42   52   61   45   55   60   43   48
## 2011 2012
##   48   56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   10   11    9   20   30   22   30   34   34   26   31   27   20   19
## 2011 2012
##   27   26

```



```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    9    9    8   20   25   18   26   31   30   22   25   24   16   18
## 2011 2012
##   24   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 = 18, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7

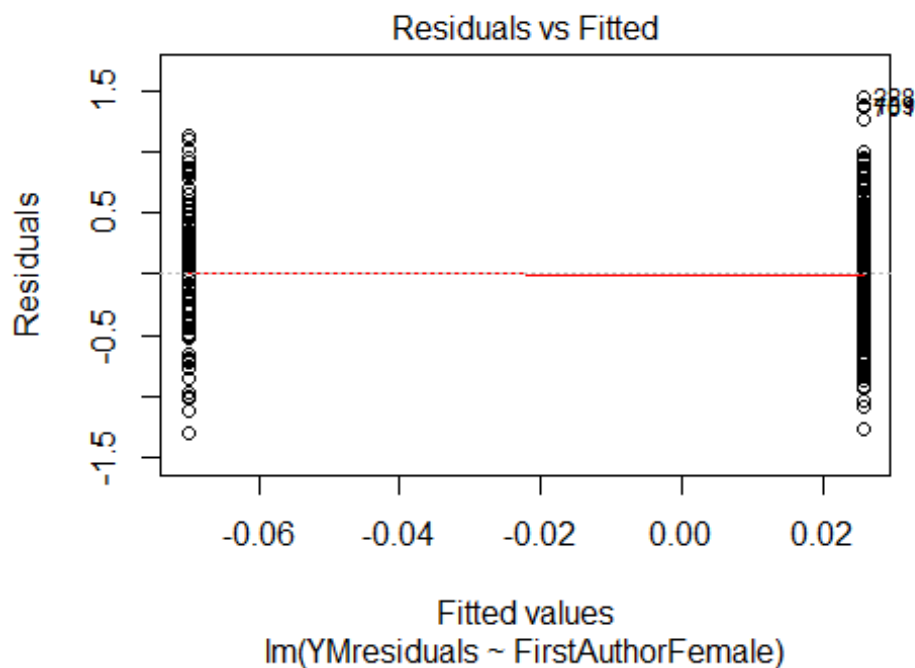
## [1] "Female first author team size 2018 geometric mean: 3.54400628740746"
## [1] "Male first author team size 2018 geometric mean: 3.36324967664711"

## 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 = 300, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.77668326129269"
## [1] "Male last author team size 2018 geometric mean: 3.25901713890948"

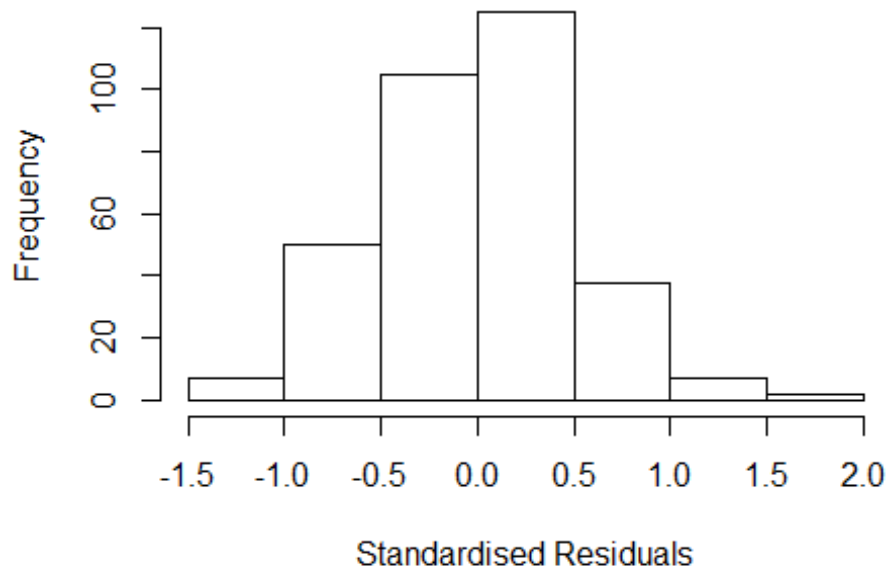
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 330, p-value = 0.6
## 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	1.378	1	1.174
LastAuthorFemale	1.332	1	1.154
UniqueAuthors	3.160	4	1.155
Year	4.083	16	1.045

Residuals from first and last author and team size



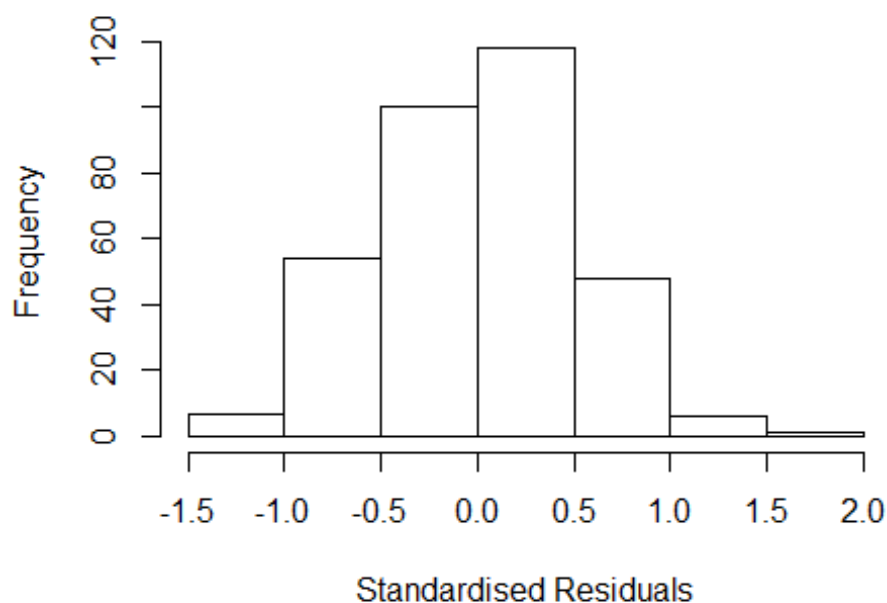
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3618 -0.3444 0.0159 0.3270 1.7130
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45323 0.30812 4.72 3.6e-06 ***
## FirstAuthorFemale1 -0.08498 0.06846 -1.24 0.215
## LastAuthorFemale1 -0.00641 0.07471 -0.09 0.932
## UniqueAuthors2 -0.15287 0.17163 -0.89 0.374
## UniqueAuthors3 0.07002 0.17213 0.41 0.684
## UniqueAuthors4 -0.03482 0.16525 -0.21 0.833
## UniqueAuthors5 0.30857 0.16938 1.82 0.069 .
## Year1997 -0.11252 0.28616 -0.39 0.694
## Year1998 -0.19331 0.30472 -0.63 0.526
## Year1999 -0.58936 0.31765 -1.86 0.064 .
```

```

## Year2000      -0.81156    0.38211   -2.12    0.034 *
## Year2001      -0.70940    0.30028   -2.36    0.019 *
## Year2002      -0.45374    0.30064   -1.51    0.132
## Year2003      -0.66219    0.27990   -2.37    0.019 *
## Year2004      -0.64304    0.27349   -2.35    0.019 *
## Year2005      -0.49644    0.27216   -1.82    0.069 .
## Year2006      -0.46177    0.29031   -1.59    0.113
## Year2007      -0.36639    0.28229   -1.30    0.195
## Year2008      -0.35396    0.28086   -1.26    0.209
## Year2009      -0.57098    0.30273   -1.89    0.060 .
## Year2010      -0.33673    0.28505   -1.18    0.238
## Year2011      -0.60293    0.27456   -2.20    0.029 *
## Year2012      -0.45191    0.28441   -1.59    0.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.464
## Multiple R-squared:  0.195, Adjusted R-squared:  0.138
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 304 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.144  0.843  0.948  0.894  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           2.99e-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 1.330 1           1.153
## LastAuthorFemale  1.202 1           1.096
## Year              1.425 16           1.011

```

Residuals from first and last author

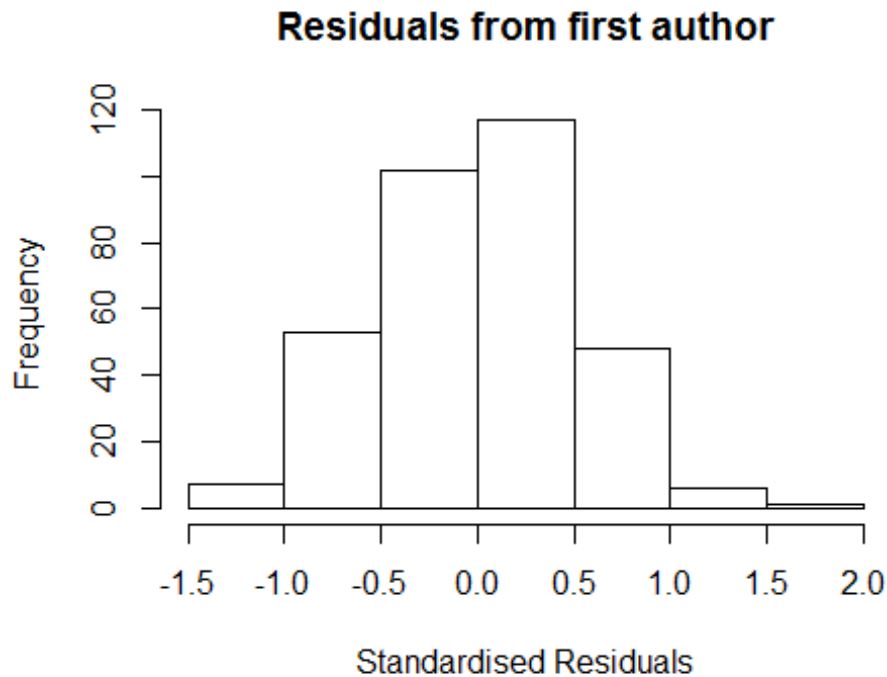


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3703 -0.3720 0.0442 0.3579 1.5110
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4275 0.2816 5.07 6.8e-07 ***
## FirstAuthorFemale1 -0.0402 0.0724 -0.56 0.579
## LastAuthorFemale1 -0.0170 0.0800 -0.21 0.832
## Year1997 -0.0946 0.3065 -0.31 0.758
## Year1998 -0.0678 0.3181 -0.21 0.831
## Year1999 -0.4390 0.3511 -1.25 0.212
## Year2000 -0.7248 0.3603 -2.01 0.045 *
## Year2001 -0.6345 0.3269 -1.94 0.053 .
## Year2002 -0.3517 0.3243 -1.08 0.279
## Year2003 -0.6402 0.2994 -2.14 0.033 *
## Year2004 -0.5346 0.2966 -1.80 0.072 .
## Year2005 -0.4144 0.2974 -1.39 0.164
```

```

## Year2006          -0.4289      0.3056   -1.40    0.161
## Year2007          -0.3186      0.3060   -1.04    0.299
## Year2008          -0.3284      0.3064   -1.07    0.285
## Year2009          -0.3478      0.3289   -1.06    0.291
## Year2010          -0.1595      0.3087   -0.52    0.606
## Year2011          -0.4733      0.3007   -1.57    0.116
## Year2012          -0.3432      0.3108   -1.10    0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0967, Adjusted R-squared:  0.0451
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 316 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.387  0.869  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      2.99e-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 1.288 1      1.135
## Year              1.288 16      1.008

```

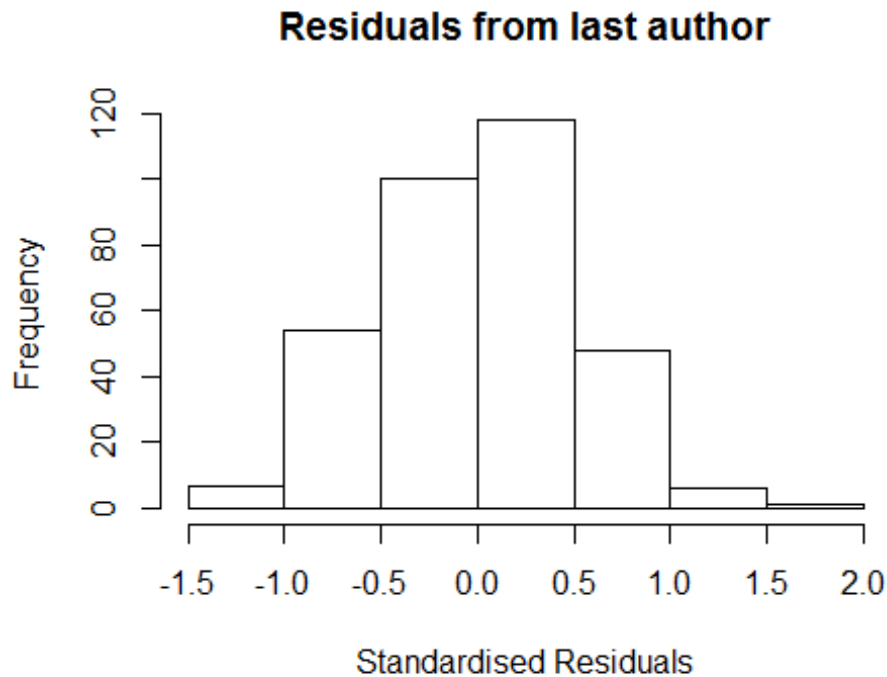


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.384 -0.368 0.043 0.356 1.513
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4267 0.2816 5.07 6.9e-07 ***
## FirstAuthorFemale1 -0.0432 0.0713 -0.61 0.545
## Year1997 -0.0934 0.3067 -0.30 0.761
## Year1998 -0.0697 0.3165 -0.22 0.826
## Year1999 -0.4416 0.3485 -1.27 0.206
## Year2000 -0.7240 0.3611 -2.00 0.046 *
## Year2001 -0.6361 0.3268 -1.95 0.052 .
## Year2002 -0.3519 0.3241 -1.09 0.278
## Year2003 -0.6414 0.2984 -2.15 0.032 *
## Year2004 -0.5350 0.2960 -1.81 0.072 .
## Year2005 -0.4144 0.2970 -1.40 0.164
## Year2006 -0.4314 0.3037 -1.42 0.156
```

```

## Year2007          -0.3222      0.3034   -1.06    0.289
## Year2008          -0.3314      0.3048   -1.09    0.278
## Year2009          -0.3470      0.3288   -1.06    0.292
## Year2010          -0.1629      0.3067   -0.53    0.596
## Year2011          -0.4736      0.3002   -1.58    0.116
## Year2012          -0.3451      0.3092   -1.12    0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0971, Adjusted R-squared:  0.0485
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 318 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.381  0.869  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      2.99e-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 1.169 1          1.081
## Year            1.169 16          1.005

```

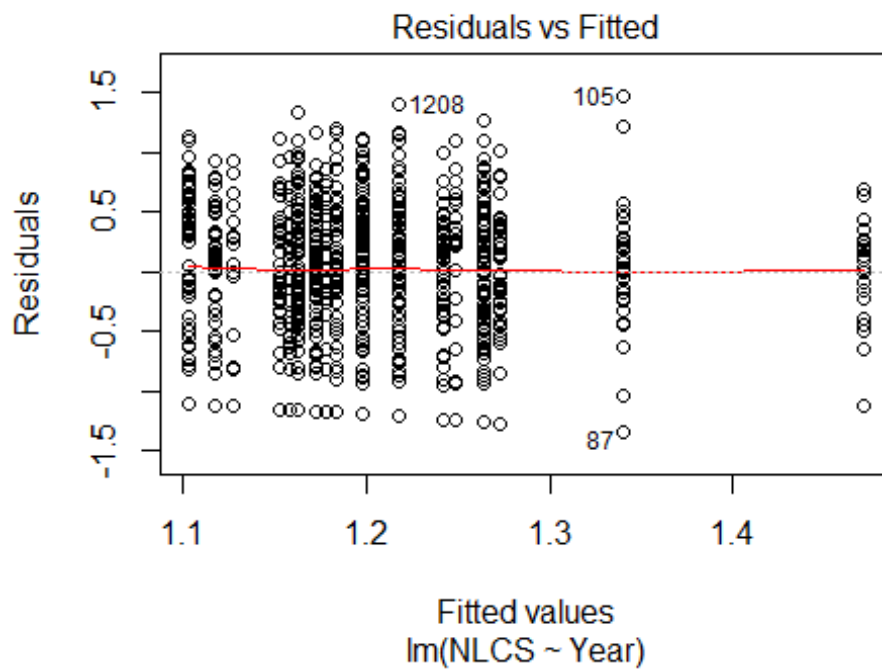
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4057 -0.3862 0.0508 0.3555 1.5180
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4331 0.2824 5.07 6.6e-07 ***
## LastAuthorFemale1 -0.0274 0.0790 -0.35 0.729
## Year1997 -0.1044 0.3055 -0.34 0.733
## Year1998 -0.0896 0.3152 -0.28 0.776
## Year1999 -0.4585 0.3487 -1.31 0.189
## Year2000 -0.7448 0.3575 -2.08 0.038 *
## Year2001 -0.6471 0.3267 -1.98 0.049 *
## Year2002 -0.3669 0.3220 -1.14 0.255
## Year2003 -0.6538 0.2985 -2.19 0.029 *
## Year2004 -0.5470 0.2951 -1.85 0.065 .
## Year2005 -0.4259 0.2963 -1.44 0.152
## Year2006 -0.4379 0.3050 -1.44 0.152
```

```

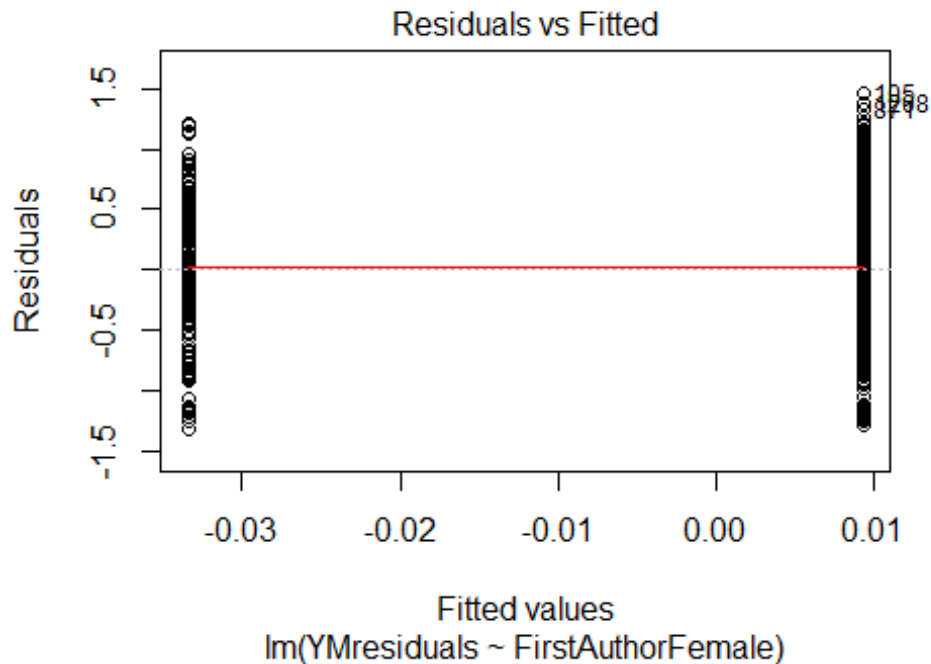
## Year2007          -0.3359      0.3039   -1.11    0.270
## Year2008          -0.3425      0.3060   -1.12    0.264
## Year2009          -0.3619      0.3287   -1.10    0.272
## Year2010          -0.1726      0.3071   -0.56    0.575
## Year2011          -0.4937      0.2956   -1.67    0.096 .
## Year2012          -0.3603      0.3069   -1.17    0.241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.523
## Multiple R-squared:  0.0967, Adjusted R-squared:  0.0481
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.380  0.872  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      2.99e-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: 334"
## [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
##   31   24   41   29   33   49   51   50   76   52   60   74   94  115  112
## 2011 2012
##  111  103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   17   34   25   22   29   46   44   55   42   50   61   82   98  106
## 2011 2012

```

```
## 97 87
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 16 32 23 20 26 43 40 51 40 44 51 75 91 100
## 2011 2012
## 94 77
## [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
```

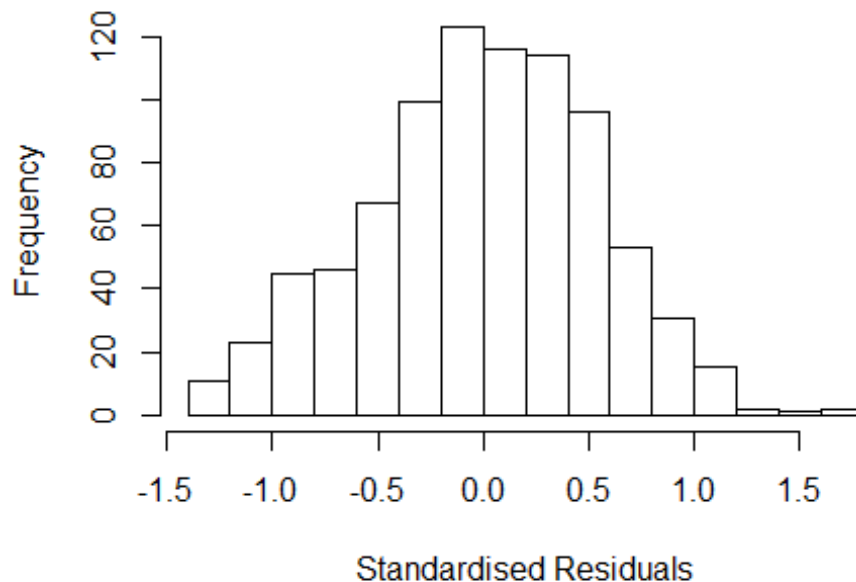


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.057, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 3.42711238966264"
## [1] "Male first author team size 2018 geometric mean: 3.11999000345608"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 800, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.2771191687868"
## [1] "Male last author team size 2018 geometric mean: 3.60494688878291"
##
## 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 1.157 1      1.076
## LastAuthorFemale  1.154 1      1.074
## UniqueAuthors     1.480 4      1.050
## Year               1.605 16     1.015
```

Residuals from first and last author and team size



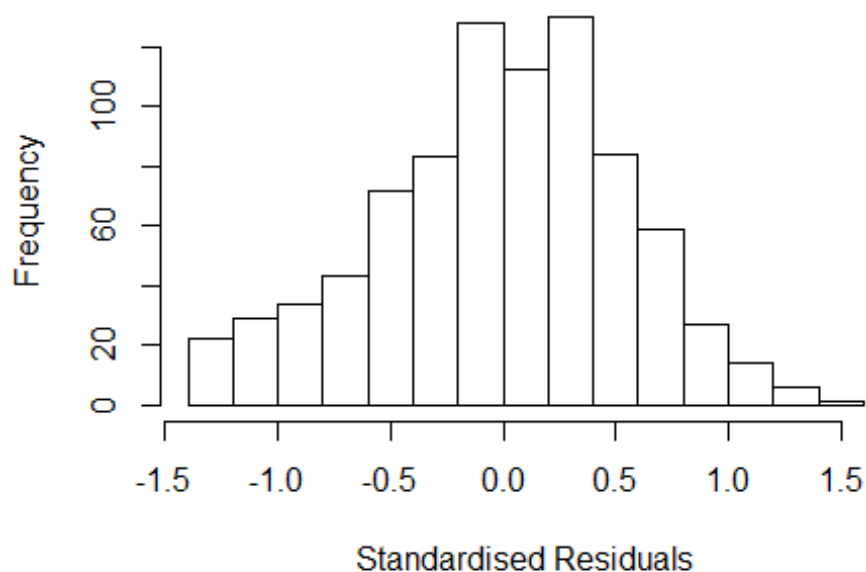
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3110 -0.3626 0.0133 0.3805 1.6322
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.188647 0.156837 7.58 9.4e-14 ***
## FirstAuthorFemale1 0.000641 0.050358 0.01 0.9899
## LastAuthorFemale1 -0.052723 0.040887 -1.29 0.1976
## UniqueAuthors2 0.235712 0.072316 3.26 0.0012 **
## UniqueAuthors3 0.300986 0.066728 4.51 7.4e-06 ***
## UniqueAuthors4 0.419129 0.074494 5.63 2.5e-08 ***
## UniqueAuthors5 0.299935 0.074260 4.04 5.9e-05 ***
## Year1997 -0.305811 0.224534 -1.36 0.1736
## Year1998 -0.076183 0.167720 -0.45 0.6498
## Year1999 -0.154657 0.183877 -0.84 0.4005
```

```

## Year2000      0.128850    0.174442    0.74    0.4603
## Year2001     -0.187972    0.174996   -1.07    0.2831
## Year2002     -0.131916    0.157638   -0.84    0.4029
## Year2003     -0.275621    0.169580   -1.63    0.1045
## Year2004     -0.248287    0.167474   -1.48    0.1386
## Year2005     -0.200149    0.166335   -1.20    0.2292
## Year2006     -0.191576    0.159420   -1.20    0.2298
## Year2007     -0.350830    0.175529   -2.00    0.0460 *
## Year2008     -0.244669    0.155966   -1.57    0.1171
## Year2009     -0.253674    0.154320   -1.64    0.1006
## Year2010     -0.183853    0.157674   -1.17    0.2439
## Year2011     -0.189314    0.152780   -1.24    0.2157
## Year2012     -0.156796    0.160035   -0.98    0.3275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.076, Adjusted R-squared:  0.0513
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 776 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.337  0.865  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      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 1.115 1      1.056
## LastAuthorFemale  1.083 1      1.040
## Year              1.183 16      1.005

```

Residuals from first and last author



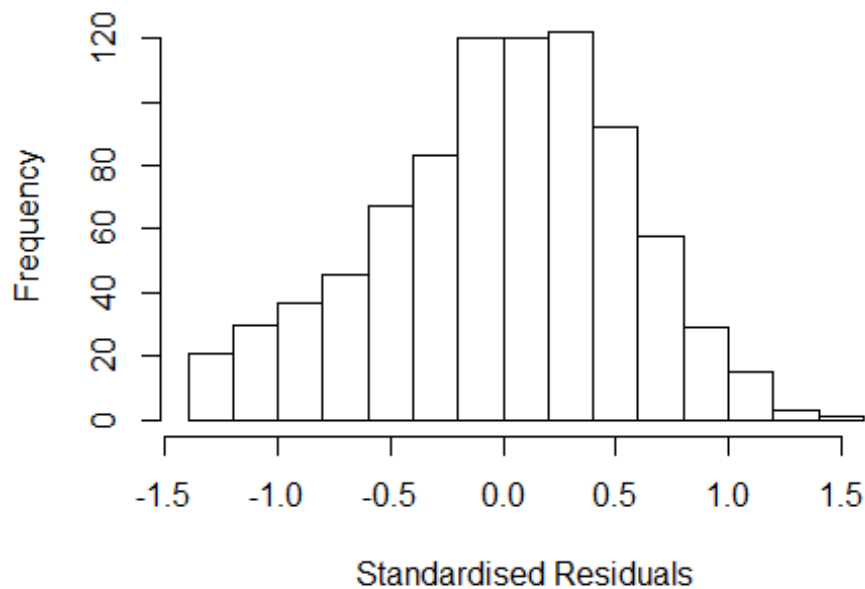
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.329 -0.386  0.024  0.365  1.430
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4387    0.1593   9.03  <2e-16 ***
## FirstAuthorFemale1  0.0497    0.0505   0.98   0.325
## LastAuthorFemale1 -0.0727    0.0404  -1.80   0.072 .
## Year1997          -0.3291    0.2349  -1.40   0.162
## Year1998          -0.1100    0.1798  -0.61   0.541
## Year1999          -0.2383    0.1974  -1.21   0.228
## Year2000           0.0681    0.1844   0.37   0.712
## Year2001          -0.2727    0.1864  -1.46   0.144
## Year2002          -0.1613    0.1707  -0.95   0.345
## Year2003          -0.3201    0.1761  -1.82   0.069 .
## Year2004          -0.2529    0.1766  -1.43   0.153
## Year2005          -0.2446    0.1748  -1.40   0.162
```

```

## Year2006          -0.2042      0.1728   -1.18    0.238
## Year2007          -0.3853      0.1984   -1.94    0.052 .
## Year2008          -0.2363      0.1692   -1.40    0.163
## Year2009          -0.2777      0.1669   -1.66    0.097 .
## Year2010          -0.1997      0.1701   -1.17    0.241
## Year2011          -0.1900      0.1670   -1.14    0.256
## Year2012          -0.1513      0.1727   -0.88    0.381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.0261, Adjusted R-squared:  0.00484
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.474  0.868  0.951  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      1.18e-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 1.1 1         1.049
## Year              1.1 16         1.003

```


Residuals from first author



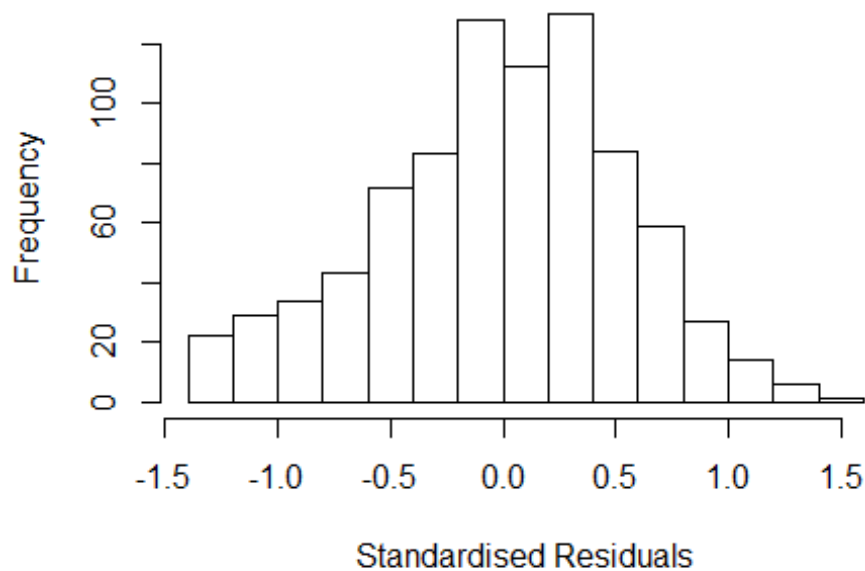
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2995 -0.3752 0.0241 0.3737 1.4772
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4093 0.1571 8.97 <2e-16 ***
## FirstAuthorFemale1 0.0313 0.0506 0.62 0.536
## Year1997 -0.3158 0.2360 -1.34 0.181
## Year1998 -0.1099 0.1788 -0.61 0.539
## Year1999 -0.2329 0.1971 -1.18 0.238
## Year2000 0.0647 0.1831 0.35 0.724
## Year2001 -0.2776 0.1866 -1.49 0.137
## Year2002 -0.1514 0.1704 -0.89 0.375
## Year2003 -0.3228 0.1753 -1.84 0.066 .
## Year2004 -0.2581 0.1768 -1.46 0.145
## Year2005 -0.2493 0.1743 -1.43 0.153
## Year2006 -0.1994 0.1724 -1.16 0.248
```

```

## Year2007          -0.3757      0.1983   -1.89    0.059 .
## Year2008          -0.2367      0.1686   -1.40    0.161
## Year2009          -0.2758      0.1659   -1.66    0.097 .
## Year2010          -0.2002      0.1697   -1.18    0.238
## Year2011          -0.1881      0.1663   -1.13    0.258
## Year2012          -0.1516      0.1718   -0.88    0.378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.00219
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 66 weights are ~= 1. The remaining 778 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.431  0.863  0.947  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      1.18e-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 1.066 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
## -1.361 -0.379  0.018  0.365  1.447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4640     0.1601   9.14 <2e-16 ***
## LastAuthorFemale1 -0.0648     0.0405  -1.60   0.110
## Year1997         -0.3103     0.2343  -1.32   0.186
## Year1998         -0.1031     0.1818  -0.57   0.571
## Year1999         -0.2351     0.2001  -1.18   0.240
## Year2000          0.0703     0.1851   0.38   0.704
## Year2001         -0.2656     0.1870  -1.42   0.156
## Year2002         -0.1581     0.1728  -0.91   0.361
## Year2003         -0.3138     0.1777  -1.77   0.078 .
## Year2004         -0.2437     0.1781  -1.37   0.172
## Year2005         -0.2336     0.1761  -1.33   0.185
## Year2006         -0.1928     0.1744  -1.11   0.269
```

```

## Year2007          -0.3719      0.1990    -1.87      0.062 .
## Year2008          -0.2214      0.1697    -1.30      0.193
## Year2009          -0.2653      0.1680    -1.58      0.115
## Year2010          -0.1917      0.1718    -1.12      0.265
## Year2011          -0.1784      0.1684    -1.06      0.290
## Year2012          -0.1406      0.1736    -0.81      0.418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.00505
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 770 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.451  0.866  0.949  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      1.18e-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: 844"

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	3049		57.00	3.846	46.332	42.4859	3.846	28.185	24.3392
##	1100	10892		51.34	22.093	51.485	29.3921	16.279	36.634	20.3546
##	1101	1277		59.44	23.077	41.026	17.9487	15.385	20.513	5.1282
##	1102	7107		38.88	22.059	30.387	8.3279	14.706	16.022	1.3162
##	1103	11043		48.45	21.831	53.994	32.1626	17.606	30.990	13.3848
##	1104	11520		53.91	19.718	41.429	21.7103	15.493	22.571	7.0785
##	1105	23235		60.70	21.818	43.455	21.6373	18.909	26.702	7.7925
##	1106	6213		39.18	15.789	62.069	46.2795	19.298	39.224	19.9259
##	1107	3104		48.61	16.667	37.500	20.8333	12.500	20.455	7.9545
##	1108	1812		32.84	37.500	26.667	-10.8333	25.000	20.000	-5.0000
##	1109	3336		52.55	25.000	43.421	18.4211	27.273	11.842	-15.4306
##	1110	10228		48.44	36.364	35.193	-1.1705	20.455	23.605	3.1506
##	1111	4461		33.15	10.811	28.148	17.3373	10.811	17.778	6.9670
##	1200	808		72.28	14.286	35.294	21.0084	14.286	44.118	29.8319
##	1201	6398		77.35	36.957	59.432	22.4750	36.957	51.938	14.9815
##	1202	6885		84.79	32.500	50.244	17.7439	32.500	46.098	13.5976
##	1203	4610		78.20	54.762	61.905	7.1429	55.952	54.762	-1.1905
##	1204	1252		78.83	16.667	61.111	44.4444	33.333	45.833	12.5000
##	1205	230		72.17	50.000	50.000	0.0000	50.000	50.000	0.0000
##	1206	288		77.78	NaN	57.692	NaN	NaN	57.692	NaN
##	1207	1750		83.66	35.714	52.000	16.2857	32.143	45.000	12.8571
##	1208	2588		84.97	53.333	58.065	4.7312	53.333	58.065	4.7312
##	1209	312		83.01	NaN	59.091	NaN	NaN	63.636	NaN
##	1210	804		86.57	25.000	62.222	37.2222	25.000	51.111	26.1111
##	1211	3697		84.28	18.868	33.333	14.4654	20.755	35.714	14.9596
##	1212	1617		86.52	21.429	38.679	17.2507	14.286	41.509	27.2237
##	1213	2286		84.82	0.000	63.736	63.7363	0.000	60.989	60.9890
##	1300	8558		65.02	21.127	53.191	32.0647	21.127	37.234	16.1073
##	1301	61		63.93	NaN	40.000	NaN	NaN	20.000	NaN
##	1302	711		66.24	42.857	43.333	0.4762	57.143	43.333	-13.8095
##	1303	14276		60.49	30.878	40.989	10.1112	11.048	24.382	13.3335
##	1304	4315		52.79	25.581	38.261	12.6795	15.116	32.174	17.0576
##	1305	4811		48.16	26.531	36.364	9.8330	18.367	24.242	5.8751
##	1306	4515		57.08	45.614	61.268	15.6536	24.561	30.986	6.4245
##	1307	8424		57.36	36.866	49.189	12.3228	21.198	27.568	6.3694
##	1308	3033		54.57	35.526	40.323	4.7963	27.632	27.419	-0.2122
##	1309	2839		56.75	34.000	69.014	35.0141	14.000	36.620	22.6197
##	1310	4609		48.58	40.000	65.455	25.4545	18.947	49.091	30.1435
##	1311	10105		53.40	34.807	48.758	13.9511	17.127	30.745	13.6183
##	1312	11534		59.88	37.838	40.491	2.6530	18.533	29.755	11.2218
##	1313	2914		61.87	50.000	44.094	-5.9055	20.833	27.559	6.7257
##	1314	8391		56.12	29.167	43.534	14.3678	20.833	32.328	11.4943
##	1315	1277		62.10	25.000	42.857	17.8571	12.500	22.857	10.3571
##	1400	2876		68.85	21.622	39.000	17.3784	10.811	32.000	21.1892
##	1401	1293		71.23	31.250	41.071	9.8214	37.500	32.143	-5.3571

## 1402	2737	66.93	17.143	32.258	15.1152	14.286	25.806	11.5207
## 1403	4425	74.10	18.333	37.569	19.2357	21.667	38.674	17.0074
## 1404	1121	58.70	26.667	13.514	-13.1532	13.333	29.730	16.3964
## 1405	2405	70.85	17.241	43.210	25.9685	20.690	38.272	17.5819
## 1406	3058	73.97	14.286	51.163	36.8771	4.762	49.612	44.8505
## 1407	3479	80.60	25.373	44.628	19.2550	23.881	42.975	19.0946
## 1408	5093	67.13	17.021	37.736	20.7146	14.894	41.038	26.1441
## 1409	3374	73.47	29.032	61.589	32.5571	16.129	62.252	46.1226
## 1410	1030	83.98	20.833	37.209	16.3760	20.833	30.233	9.3992
## 1500	6682	32.62	12.857	24.828	11.9704	1.429	9.655	8.2266
## 1501	350	31.43	0.000	26.667	26.6667	0.000	13.333	13.3333
## 1502	2583	47.12	29.167	25.758	-3.4091	20.833	19.697	-1.1364
## 1503	2082	53.99	14.583	30.645	16.0618	4.167	20.968	16.8011
## 1504	96	58.33	0.000	60.000	60.0000	100.000	20.000	-80.0000
## 1505	1268	49.37	15.385	27.778	12.3932	0.000	16.667	16.6667
## 1506	652	36.20	0.000	28.571	28.5714	0.000	28.571	28.5714
## 1507	1161	31.70	20.000	26.923	6.9231	0.000	3.846	3.8462
## 1508	559	37.75	50.000	32.258	-17.7419	0.000	19.355	19.3548
## 1600	13728	48.11	15.862	29.014	13.1520	9.310	16.901	7.5911
## 1601	419	63.25	0.000	36.364	36.3636	0.000	27.273	27.2727
## 1602	3582	53.63	23.944	41.379	17.4356	8.451	18.391	9.9401
## 1603	1800	41.56	8.696	36.000	27.3043	0.000	24.000	24.0000
## 1604	2996	59.15	23.404	29.508	6.1039	7.447	22.951	15.5040
## 1605	6299	61.15	22.222	34.058	11.8357	6.536	16.667	10.1307
## 1606	7332	52.43	22.013	31.013	9.0001	3.774	18.987	15.2138
## 1607	2713	54.99	11.538	42.308	30.7692	13.462	25.000	11.5385
## 1700	1879	48.70	38.462	31.481	-6.9801	19.231	24.074	4.8433
## 1701	225	56.44	NaN	20.000	NaN	NaN	20.000	NaN
## 1702	2462	50.20	13.043	22.222	9.1787	10.870	33.333	22.4638
## 1703	1675	47.10	15.385	16.667	1.2821	7.692	16.667	8.9744
## 1704	709	47.11	22.222	30.435	8.2126	22.222	34.783	12.5604
## 1705	2649	48.89	3.704	24.390	20.6865	0.000	26.829	26.8293
## 1706	7382	45.25	15.873	26.531	10.6576	9.524	19.388	9.8639
## 1707	1045	42.01	16.667	21.053	4.3860	16.667	10.526	-6.1404
## 1708	1400	40.79	20.000	11.538	-8.4615	10.000	19.231	9.2308
## 1709	1236	60.52	25.000	34.921	9.9206	12.500	36.508	24.0079
## 1710	2793	52.95	15.625	28.571	12.9464	9.375	22.619	13.2440
## 1711	1722	36.59	8.696	10.811	2.1152	8.696	10.811	2.1152
## 1712	4738	43.39	7.812	17.708	9.8958	6.250	23.958	17.7083
## 1800	645	73.18	0.000	48.387	48.3871	0.000	45.161	45.1613
## 1801	19	73.68	NaN	100.000	NaN	NaN	0.000	NaN
## 1802	1248	54.17	21.429	18.605	-2.8239	21.429	23.256	1.8272
## 1803	1720	54.07	12.500	25.926	13.4259	12.500	14.815	2.3148
## 1804	1599	56.04	3.704	11.628	7.9242	3.704	11.628	7.9242
## 1900	2933	32.94	22.222	24.242	2.0202	18.519	22.727	4.2088
## 1901	3272	29.55	2.703	33.962	31.2596	5.405	13.208	7.8021
## 1902	3498	48.31	11.628	48.182	36.5539	11.628	28.182	16.5539
## 1903	649	51.31	0.000	40.000	40.0000	0.000	40.000	40.0000
## 1904	3738	55.54	12.500	47.403	34.9026	18.750	35.714	16.9643
## 1905	697	54.95	0.000	38.462	38.4615	9.091	26.923	17.8322

## 1906	4749	46.26	6.452	29.126	22.6746	9.677	13.592	3.9148
## 1907	4160	49.66	4.082	36.441	32.3590	4.082	22.881	18.7997
## 1908	3470	49.74	4.348	32.143	27.7950	13.043	10.714	-2.3292
## 1909	4091	28.48	18.182	15.672	-2.5102	0.000	14.179	14.1791
## 1910	4832	58.03	18.333	45.562	27.2288	15.000	23.077	8.0769
## 1911	2148	57.45	18.750	54.902	36.1520	22.917	11.765	-11.1520
## 1912	3044	39.88	11.905	24.194	12.2888	14.286	6.452	-7.8341
## 1913	472	51.06	14.286	29.412	15.1261	14.286	29.412	15.1261
## 2000	1541	66.13	16.667	20.000	3.3333	8.333	26.154	17.8205
## 2001	1099	68.06	35.714	31.343	-4.3710	28.571	23.881	-4.6908
## 2002	10050	67.03	11.189	28.871	17.6826	7.692	24.147	16.4547
## 2003	3866	61.72	21.875	22.143	0.2679	12.500	16.429	3.9286
## 2100	1567	45.69	0.000	25.000	25.0000	0.000	21.429	21.4286
## 2101	275	44.00	NaN	12.500	NaN	NaN	18.750	NaN
## 2102	2895	31.40	10.000	18.824	8.8235	5.000	16.471	11.4706
## 2103	2010	31.84	10.526	20.690	10.1633	5.263	24.138	18.8748
## 2104	572	35.14	0.000	35.484	35.4839	0.000	32.258	32.2581
## 2105	3597	40.98	5.556	29.670	24.1148	0.000	22.527	22.5275
## 2200	2946	44.16	22.388	40.476	18.0881	4.478	19.048	14.5700
## 2201	702	58.55	0.000	50.000	50.0000	0.000	38.889	38.8889
## 2202	1464	39.28	14.286	11.111	-3.1746	7.143	2.778	-4.3651
## 2203	683	41.14	0.000	40.000	40.0000	0.000	26.667	26.6667
## 2204	3221	42.75	41.176	30.075	-11.1013	29.412	19.549	-9.8629
## 2205	5947	32.49	20.000	20.362	0.3620	3.333	14.027	10.6938
## 2206	709	24.40	25.000	8.333	-16.6667	25.000	0.000	-25.0000
## 2207	3601	35.63	8.108	6.667	-1.4414	8.108	4.444	-3.6637
## 2208	9509	33.78	10.976	11.111	0.1355	7.317	9.091	1.7738
## 2209	4589	34.84	0.000	22.481	22.4806	33.333	17.054	-16.2791
## 2210	9573	24.99	13.462	15.026	1.5644	3.846	10.881	7.0347
## 2211	7166	24.23	14.286	14.286	0.0000	4.762	6.349	1.5873
## 2212	1222	39.12	14.286	30.612	16.3265	14.286	12.245	-2.0408
## 2213	2175	47.95	14.286	34.831	20.5457	28.571	29.213	0.6421
## 2214	440	55.91	10.000	21.429	11.4286	20.000	28.571	8.5714
## 2215	3403	36.23	14.286	21.053	6.7669	0.000	15.132	15.1316
## 2216	783	61.17	0.000	36.364	36.3636	0.000	27.273	27.2727
## 2300	5671	54.81	25.000	34.359	9.3590	26.190	27.179	0.9890
## 2301	1250	68.96	28.571	56.842	28.2707	21.429	47.368	25.9398
## 2302	1501	48.17	6.250	35.417	29.1667	6.250	39.583	33.3333
## 2303	12896	60.07	17.986	41.889	23.9035	17.986	27.926	9.9405
## 2304	4886	45.19	19.048	37.989	18.9412	14.286	22.346	8.0607
## 2305	3274	43.16	13.333	31.613	18.2796	0.000	24.516	24.5161
## 2306	1604	64.78	7.143	43.158	36.0150	14.286	27.368	13.0827
## 2307	2185	54.42	20.000	61.491	41.4907	26.667	39.130	12.4638
## 2308	8114	65.94	30.864	47.230	16.3654	32.099	34.037	1.9382
## 2309	4688	66.53	15.789	46.535	30.7452	13.158	30.198	17.0401
## 2310	3851	46.43	11.111	38.068	26.9571	0.000	28.977	28.9773
## 2311	2411	45.00	7.143	36.752	29.6093	0.000	29.060	29.0598
## 2312	6377	45.05	5.556	33.862	28.3069	1.389	18.519	17.1296
## 2400	1754	57.92	27.778	49.398	21.6198	11.111	36.145	25.0335
## 2401	26	65.38	NaN	50.000	NaN	NaN	50.000	NaN

##	2402	2080	44.81	43.333	50.000	6.6667	23.333	24.000	0.6667
##	2403	7056	55.14	38.514	59.659	21.1456	16.216	35.795	19.5792
##	2404	4165	56.71	49.398	60.811	11.4132	18.072	29.054	10.9818
##	2405	2292	49.65	50.980	47.761	-3.2192	13.725	23.881	10.1551
##	2406	2117	57.16	37.500	67.123	29.6233	15.000	32.877	17.8767
##	2500	10916	30.91	13.333	21.245	7.9121	3.333	20.147	16.8132
##	2501	350	37.71	25.000	54.545	29.5455	25.000	45.455	20.4545
##	2502	2188	45.89	27.273	24.691	-2.5814	18.182	16.049	-2.1324
##	2503	3121	31.05	10.256	23.404	13.1478	7.692	17.021	9.3290
##	2504	6226	32.93	10.000	20.482	10.4819	5.000	16.867	11.8675
##	2505	7255	40.26	14.286	26.154	11.8681	6.250	23.077	16.8269
##	2506	3743	30.32	11.429	24.286	12.8571	8.571	18.571	10.0000
##	2507	3549	37.93	3.922	32.258	28.3365	1.961	12.903	10.9424
##	2508	4323	36.80	13.208	17.778	4.5702	3.774	20.000	16.2264
##	2600	2594	52.89	19.565	22.222	2.6570	17.391	16.667	-0.7246
##	2601	641	46.96	0.000	27.778	27.7778	0.000	16.667	16.6667
##	2602	859	54.83	0.000	6.897	6.8966	0.000	10.345	10.3448
##	2603	825	42.91	18.182	11.538	-6.6434	9.091	3.846	-5.2448
##	2604	5521	40.86	13.636	12.097	-1.5396	13.636	12.097	-1.5396
##	2605	1362	39.94	18.182	13.793	-4.3887	27.273	3.448	-23.8245
##	2606	686	35.57	NaN	11.429	NaN	NaN	11.429	NaN
##	2607	773	51.36	9.091	10.526	1.4354	18.182	21.053	2.8708
##	2608	489	59.92	25.000	7.692	-17.3077	37.500	15.385	-22.1154
##	2609	179	72.63	0.000	0.000	0.0000	0.000	0.000	0.0000
##	2610	1305	37.55	16.667	10.345	-6.3218	16.667	10.345	-6.3218
##	2611	3036	36.43	12.500	21.622	9.1216	12.500	20.270	7.7703
##	2612	488	38.32	0.000	0.000	0.0000	0.000	0.000	0.0000
##	2613	3556	52.22	8.772	15.385	6.6127	5.263	11.538	6.2753
##	2614	1628	47.42	10.345	3.704	-6.6411	13.793	11.111	-2.6820
##	2700	19885	67.70	34.969	57.233	22.2634	27.607	42.977	15.3696
##	2701	6290	62.61	33.708	66.975	33.2674	24.719	46.914	22.1945
##	2702	910	54.29	40.000	32.432	-7.5676	13.333	37.838	24.5045
##	2703	3150	42.98	26.667	51.402	24.7352	33.333	40.187	6.8536
##	2704	1028	47.86	33.333	20.000	-13.3333	23.810	33.333	9.5238
##	2705	6757	66.18	16.418	31.633	15.2147	5.970	21.429	15.4584
##	2706	3044	54.60	30.769	50.000	19.2308	7.692	31.522	23.8294
##	2707	1238	68.82	75.000	47.541	-27.4590	0.000	44.262	44.2623
##	2708	2421	59.40	36.585	57.143	20.5575	21.951	34.694	12.7427
##	2709	20	65.00	NaN	100.000	NaN	NaN	100.000	NaN
##	2710	569	47.45	0.000	88.889	88.8889	0.000	55.556	55.5556
##	2711	2747	72.01	0.000	53.623	53.6232	0.000	32.609	32.6087
##	2712	4933	48.71	32.812	63.636	30.8239	10.938	41.212	30.2746
##	2713	2562	60.62	43.333	58.491	15.1572	23.333	50.000	26.6667
##	2714	3056	72.97	100.000	67.273	-32.7273	60.000	47.273	-12.7273
##	2715	2715	50.98	32.258	30.108	-2.1505	25.806	22.581	-3.2258
##	2716	2336	54.32	40.909	63.415	22.5055	9.091	36.585	27.4945
##	2717	2902	69.37	42.308	64.789	22.4810	57.692	56.338	-1.3543
##	2718	1552	66.56	33.333	61.314	27.9805	33.333	45.255	11.9221
##	2719	6168	74.72	55.000	68.101	13.1013	50.000	57.468	7.4684
##	2720	2706	49.78	28.125	63.158	35.0329	25.000	32.895	7.8947

##	2721	1324	57.02	25.806	27.500	1.6935	19.355	22.500	3.1452
##	2722	1105	50.50	58.333	48.571	-9.7619	25.000	37.143	12.1429
##	2723	3862	54.17	31.481	66.087	34.6055	14.815	42.609	27.7939
##	2724	4014	34.45	26.471	58.434	31.9631	14.706	30.723	16.0170
##	2725	6298	57.00	54.902	54.035	-0.8669	7.843	32.281	24.4376
##	2726	2512	54.10	50.000	50.435	0.4348	15.000	31.304	16.3043
##	2727	1690	68.28	37.500	47.541	10.0410	18.750	29.508	10.7582
##	2728	8574	62.89	30.252	48.148	17.8960	25.210	38.462	13.2515
##	2729	6485	58.26	34.737	79.655	44.9183	31.579	67.241	35.6624
##	2730	7099	60.52	35.088	43.532	8.4441	24.561	25.051	0.4899
##	2731	4717	57.94	37.500	47.967	10.4675	22.222	27.642	5.4201
##	2732	8019	65.46	22.059	38.663	16.6047	13.235	27.208	13.9723
##	2733	2410	62.49	27.660	58.025	30.3651	23.404	41.975	18.5711
##	2734	4304	69.59	37.864	53.478	15.6142	30.097	37.391	7.2942
##	2735	7719	60.53	38.806	64.781	25.9755	26.866	51.671	24.8053
##	2736	4590	65.80	28.261	60.000	31.7391	21.739	43.913	22.1739
##	2737	4818	62.66	27.835	33.889	6.0538	16.495	36.111	19.6163
##	2738	13770	71.76	33.493	61.898	28.4050	37.799	45.401	7.6024
##	2739	16922	66.52	52.198	68.357	16.1594	37.363	53.473	16.1104
##	2740	4242	61.93	28.125	45.161	17.0363	15.625	34.194	18.5685
##	2741	5967	51.62	16.393	22.527	6.1340	19.672	15.385	-4.2875
##	2742	3377	69.53	52.632	76.296	23.6647	47.368	62.222	14.8538
##	2743	2057	44.09	44.444	82.258	37.8136	22.222	59.677	37.4552
##	2744	5	80.00	NaN	NaN	NaN	NaN	NaN	NaN
##	2745	1198	50.00	12.500	48.571	36.0714	12.500	28.571	16.0714
##	2746	9347	56.91	10.084	32.494	22.4102	9.244	17.391	8.1476
##	2747	1391	53.99	21.429	41.379	19.9507	21.429	20.690	-0.7389
##	2748	1418	55.57	8.333	35.714	27.3810	8.333	24.286	15.9524
##	2800	5068	58.07	38.554	46.970	8.4155	21.687	33.838	12.1516
##	2801	426	68.08	NaN	72.000	NaN	NaN	52.000	NaN
##	2802	2117	70.05	32.143	75.000	42.8571	50.000	40.517	-9.4828
##	2803	1346	69.76	26.667	61.053	34.3860	26.667	42.105	15.4386
##	2804	2958	54.94	48.718	54.286	5.5678	28.205	33.333	5.1282
##	2805	1985	69.57	51.724	64.545	12.8213	37.931	43.636	5.7053
##	2806	801	71.41	66.667	67.742	1.0753	50.000	48.387	-1.6129
##	2807	510	44.51	33.333	77.778	44.4444	50.000	44.444	-5.5556
##	2808	4915	60.04	36.735	48.087	11.3527	24.490	37.158	12.6687
##	2809	2554	58.73	38.235	50.667	12.4314	29.412	26.667	-2.7451
##	2900	5554	74.25	65.517	78.827	13.3101	67.241	73.941	6.7000
##	2901	200	56.00	NaN	93.333	NaN	NaN	40.000	NaN
##	2902	1371	70.02	18.182	80.357	62.1753	9.091	50.000	40.9091
##	2903	30	76.67	NaN	66.667	NaN	NaN	66.667	NaN
##	2904	39	87.18	NaN	66.667	NaN	NaN	33.333	NaN
##	2905	1277	77.92	NaN	74.359	NaN	NaN	60.256	NaN
##	2906	493	82.35	80.000	70.370	-9.6296	60.000	70.370	10.3704
##	2907	1033	81.12	50.000	62.069	12.0690	33.333	55.172	21.8391
##	2908	143	53.15	NaN	85.714	NaN	NaN	42.857	NaN
##	2909	738	67.62	33.333	77.551	44.2177	33.333	67.347	34.0136
##	2910	668	81.59	42.857	53.846	10.9890	42.857	43.590	0.7326
##	2911	662	70.09	41.176	80.000	38.8235	47.059	73.333	26.2745

##	2912	579	83.25	66.667	90.909	24.2424	66.667	69.697	3.0303
##	2913	857	77.60	NaN	97.260	NaN	NaN	90.411	NaN
##	2914	312	75.00	33.333	100.000	66.6667	33.333	61.538	28.2051
##	2915	2	100.00	NaN	0.000	NaN	NaN	0.000	NaN
##	2916	4356	62.44	41.176	83.577	42.4002	23.529	63.869	40.3392
##	2917	233	77.68	100.000	90.000	-10.0000	100.000	60.000	-40.0000
##	2918	0	NA	NA	NA	NA	NA	NA	NA
##	2919	419	79.71	66.667	90.323	23.6559	100.000	87.097	-12.9032
##	2920	94	76.60	NaN	75.000	NaN	NaN	50.000	NaN
##	2921	1274	72.06	0.000	59.574	59.5745	0.000	60.638	60.6383
##	2922	308	87.01	NaN	100.000	NaN	NaN	66.667	NaN
##	2923	179	82.12	NaN	84.615	NaN	NaN	69.231	NaN
##	3000	560	65.36	33.333	69.231	35.8974	50.000	34.615	-15.3846
##	3001	68	72.06	NaN	66.667	NaN	NaN	50.000	NaN
##	3002	2527	66.60	32.653	30.612	-2.0408	8.163	15.306	7.1429
##	3003	2278	60.89	50.000	53.086	3.0864	19.231	43.210	23.9791
##	3004	6015	59.95	33.696	52.577	18.8817	20.109	38.660	18.5511
##	3005	2149	61.80	10.714	54.545	43.8312	39.286	36.364	-2.9221
##	3100	6119	37.62	0.000	22.727	22.7273	0.000	12.338	12.3377
##	3101	1839	23.82	0.000	30.769	30.7692	20.000	7.692	-12.3077
##	3102	1334	39.28	14.286	32.000	17.7143	0.000	28.000	28.0000
##	3103	2064	36.39	11.111	17.391	6.2802	14.815	10.870	-3.9452
##	3104	12720	32.13	10.476	15.464	4.9877	7.619	11.856	4.2366
##	3105	2351	38.45	16.667	19.672	3.0055	11.111	21.311	10.2004
##	3106	1764	33.22	0.000	15.000	15.0000	0.000	5.000	5.0000
##	3107	5739	33.84	4.762	19.192	14.4300	1.587	9.091	7.5036
##	3108	850	48.94	14.286	21.429	7.1429	0.000	25.000	25.0000
##	3109	1715	35.34	13.636	8.571	-5.0649	13.636	8.571	-5.0649
##	3110	2017	35.30	13.043	15.789	2.7460	0.000	21.053	21.0526
##	3200	4917	79.36	42.574	55.872	13.2976	45.545	48.043	2.4982
##	3201	1095	77.99	47.619	60.000	12.3810	38.095	52.500	14.4048
##	3202	3212	74.47	65.000	59.314	-5.6863	45.000	48.529	3.5294
##	3203	5180	76.93	37.333	60.194	22.8608	42.667	52.427	9.7605
##	3204	6450	78.31	49.091	70.305	21.2137	50.000	58.376	8.3756
##	3205	4039	78.39	36.585	54.585	17.9998	40.244	44.105	3.8609
##	3206	1685	76.74	58.824	60.825	2.0012	70.588	51.546	-19.0418
##	3207	4219	78.41	53.704	65.690	11.9867	55.556	61.088	5.5323
##	3300	2935	76.46	20.690	45.198	24.5081	17.241	44.633	27.3914
##	3301	3643	78.26	28.814	65.854	37.0401	28.814	62.195	33.3816
##	3302	1314	77.93	0.000	58.824	58.8235	0.000	47.059	47.0588
##	3303	4205	72.53	25.000	61.146	36.1465	26.471	49.682	23.2109
##	3304	22028	79.19	40.072	66.106	26.0336	36.101	60.657	24.5560
##	3305	12046	70.89	25.143	51.699	26.5566	24.571	49.732	25.1602
##	3306	5789	74.81	42.857	71.389	28.5317	39.286	60.000	20.7143
##	3307	1339	70.35	23.810	41.270	17.4603	23.810	38.095	14.2857
##	3308	7720	79.95	44.660	53.965	9.3046	47.573	48.678	1.1056
##	3309	2793	78.02	28.261	59.091	30.8300	26.087	51.136	25.0494
##	3310	5135	78.42	55.319	62.551	7.2323	56.383	58.436	2.0532
##	3311	1375	66.55	25.000	34.783	9.7826	50.000	30.435	-19.5652
##	3312	15078	81.65	30.645	55.937	25.2923	32.796	50.208	17.4126

```

## 3313 1662      61.43   8.333  32.692  24.3590   8.333  24.038  15.7051
## 3314 3395      80.18  33.962  58.796  24.8340  32.075  54.167  22.0912
## 3315 3314      83.55  48.000  60.648  12.6481  52.000  59.259   7.2593
## 3316 7128      82.76  27.586  54.438  26.8517  32.759  54.043  21.2848
## 3317 1071      74.79  40.909  55.172  14.2633  40.909  63.793  22.8840
## 3318 1696      84.96  90.323  78.899 -11.4235  90.323  71.560 -18.7629
## 3319 1095      84.02  44.444  61.765  17.3203  44.444  67.647  23.2026
## 3320 4790      80.25  21.875  36.913  15.0378  21.875  35.906  14.0310
## 3321 1939      81.79  20.833  55.652  34.8188  29.167  48.696  19.5290
## 3322 1740      76.84  14.286  51.587  37.3016  14.286  46.032  31.7460
## 3400 4463      31.59  33.333  70.922  37.5887  23.333  41.135  17.8014
## 3401  174      31.61   0.000  50.000  50.0000   0.000  20.000  20.0000
## 3402  229      19.21 100.000  40.000 -60.0000 100.000  40.000 -60.0000
## 3403  777      34.11  25.000  66.667  41.6667  25.000  30.556   5.5556
## 3404  439      35.31  20.000  50.000  30.0000  20.000  25.000   5.0000
## 3500 1742      28.24  20.000  39.130  19.1304  20.000  17.391  -2.6087
## 3501   67      20.90   NaN     NaN     NaN     NaN     NaN     NaN
## 3502   0        NA     NA     NA     NA     NA     NA     NA
## 3503   1        0.00   NaN     NaN     NaN     NaN     NaN     NaN
## 3504  385      48.57  50.000  23.077 -26.9231  16.667   7.692  -8.9744
## 3505   86      32.56   NaN  66.667     NaN     NaN  33.333     NaN
## 3506   87      36.78 100.000 100.000   0.0000   0.000   0.000   0.0000
## 3600  537      68.72  25.000  77.778  52.7778  50.000 100.000  50.0000
## 3601  573      84.29   NaN  66.667     NaN     NaN  66.667     NaN
## 3602  198      76.77   NaN  37.500     NaN     NaN  31.250     NaN
## 3603  244      70.49 100.000  66.667 -33.3333  66.667  50.000 -16.6667
## 3604  661      61.42   NaN  66.667     NaN     NaN  59.259     NaN
## 3605  238      53.36   NaN  38.235     NaN     NaN  32.353     NaN
## 3606   6     100.00   NaN     NaN     NaN     NaN     NaN     NaN
## 3607  424      47.64 100.000  33.333 -66.6667   0.000  11.111  11.1111
## 3608   8     100.00   NaN 100.000     NaN     NaN 100.000     NaN
## 3609  911      78.81  83.333  85.185   1.8519 100.000  81.481 -18.5185
## 3610  800      76.12  26.316  53.333  27.0175  15.789  36.667  20.8772
## 3611  996      76.71 100.000  72.500 -27.5000 100.000  61.250 -38.7500
## 3612 6441      71.90  40.506  53.186  12.6793  39.241  40.166   0.9257
## 3613  136      80.15   0.000  33.333  33.3333   0.000  33.333  33.3333
## 3614 1276      47.57  14.286  30.189  15.9030  28.571  32.075   3.5040
## 3615   0        NA     NA     NA     NA     NA     NA     NA
## 3616 1842      77.42  57.143  84.906  27.7628  57.143  67.925  10.7817

```

```

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         584  27.740   3.846  34.694  30.8477  15.582   3.846  20.408
## 1100        2408  34.593  22.093  44.186  22.0930  21.429  16.279  24.313
## 1101         504  27.183  23.077  20.930  -2.1467  19.048  15.385  18.605

```

## 1102	1718	29.453	22.059	30.675	8.6160	19.383	14.706	20.859
## 1103	3387	36.522	21.831	41.096	19.2649	23.354	17.606	22.945
## 1104	4027	30.718	19.718	32.544	12.8261	21.182	15.493	24.852
## 1105	8988	34.112	21.818	37.821	16.0031	22.563	18.909	22.399
## 1106	1330	37.444	15.789	45.763	29.9732	23.609	19.298	26.271
## 1107	934	23.555	16.667	18.182	1.5152	15.846	12.500	22.727
## 1108	422	31.754	37.500	16.129	-21.3710	25.829	25.000	25.806
## 1109	1230	30.163	25.000	30.000	5.0000	21.138	27.273	20.000
## 1110	3297	35.608	36.364	33.465	-2.8991	22.050	20.455	16.929
## 1111	780	26.667	10.811	25.000	14.1892	13.718	10.811	19.737
## 1200	298	48.658	14.286	60.000	45.7143	45.302	14.286	52.000
## 1201	2738	48.977	36.957	53.819	16.8629	44.010	36.957	47.569
## 1202	3131	40.307	32.500	43.979	11.4791	38.997	32.500	44.241
## 1203	2177	55.949	54.762	56.604	1.8419	51.677	55.952	48.113
## 1204	528	31.061	16.667	42.105	25.4386	27.462	33.333	45.614
## 1205	96	22.917	50.000	31.579	-18.4211	23.958	50.000	36.842
## 1206	110	59.091	NaN	42.857	NaN	54.545	NaN	42.857
## 1207	837	38.829	35.714	47.423	11.7084	37.276	32.143	51.546
## 1208	1070	50.187	53.333	48.667	-4.6667	49.346	53.333	50.000
## 1209	158	47.468	NaN	36.364	NaN	46.203	NaN	40.909
## 1210	348	48.276	25.000	55.932	30.9322	43.966	25.000	49.153
## 1211	1706	27.433	18.868	23.656	4.7880	27.315	20.755	24.194
## 1212	711	29.817	21.429	32.967	11.5385	28.833	14.286	30.769
## 1213	887	48.816	0.000	50.323	50.3226	47.802	0.000	45.806
## 1300	2010	38.159	21.127	47.610	26.4828	23.085	21.127	25.498
## 1301	10	20.000	NaN	0.000	NaN	0.000	NaN	0.000
## 1302	284	58.099	42.857	67.857	25.0000	36.972	57.143	32.143
## 1303	6231	37.570	30.878	40.045	9.1671	15.311	11.048	21.267
## 1304	1433	36.636	25.581	50.442	24.8611	17.167	15.116	20.354
## 1305	1467	39.468	26.531	51.449	24.9187	20.723	18.367	29.710
## 1306	1626	50.923	45.614	55.696	10.0822	27.860	24.561	33.544
## 1307	3512	48.035	36.866	55.844	18.9778	21.754	21.198	23.810
## 1308	1152	42.708	35.526	43.750	8.2237	21.441	27.632	20.833
## 1309	1100	53.727	34.000	58.537	24.5366	28.364	14.000	29.268
## 1310	1520	50.855	40.000	61.947	21.9469	24.868	18.947	31.858
## 1311	3449	44.709	34.807	51.667	16.8600	23.398	17.127	27.000
## 1312	4746	43.911	37.838	48.876	11.0386	19.195	18.533	24.719
## 1313	1089	41.414	50.000	35.714	-14.2857	19.284	20.833	23.810
## 1314	3190	39.091	29.167	45.228	16.0615	21.473	20.833	23.237
## 1315	583	31.046	25.000	35.294	10.2941	16.123	12.500	23.529
## 1400	1223	32.052	21.622	36.364	14.7420	28.291	10.811	36.364
## 1401	619	32.633	31.250	40.909	9.6591	31.502	37.500	29.545
## 1402	1190	29.076	17.143	39.286	22.1429	26.975	14.286	32.143
## 1403	2117	32.877	18.333	41.709	23.3752	31.318	21.667	43.216
## 1404	430	34.419	26.667	26.667	0.0000	28.372	13.333	30.000
## 1405	1096	33.394	17.241	40.426	23.1842	29.653	20.690	39.362
## 1406	1375	38.036	14.286	38.667	24.3810	35.564	4.762	43.333
## 1407	1793	41.885	25.373	44.048	18.6745	40.100	23.881	38.690
## 1408	2041	35.767	17.021	38.496	21.4743	31.602	14.894	36.283
## 1409	1570	45.924	29.032	47.531	18.4986	41.847	16.129	45.679

## 1410	563	37.300	20.833	56.250	35.4167	37.123	20.833	54.167
## 1500	1161	16.451	12.857	16.260	3.4030	10.853	1.429	15.447
## 1501	55	14.545	0.000	0.000	0.0000	3.636	0.000	0.000
## 1502	661	34.644	29.167	37.079	7.9120	16.944	20.833	15.730
## 1503	728	22.665	14.583	30.357	15.7738	9.890	4.167	17.857
## 1504	42	26.190	0.000	50.000	50.0000	26.190	100.000	0.000
## 1505	486	26.543	15.385	40.000	24.6154	9.259	0.000	8.000
## 1506	141	29.787	0.000	36.364	36.3636	24.113	0.000	36.364
## 1507	236	14.407	20.000	23.077	3.0769	13.983	0.000	7.692
## 1508	90	22.222	50.000	16.667	-33.3333	15.556	0.000	33.333
## 1600	4329	22.292	15.862	25.895	10.0332	13.213	9.310	17.080
## 1601	167	18.563	0.000	20.000	20.0000	15.569	0.000	40.000
## 1602	1289	26.299	23.944	31.818	7.8745	14.197	8.451	16.364
## 1603	482	25.104	8.696	30.769	22.0736	13.485	0.000	17.308
## 1604	1327	24.115	23.404	25.610	2.2055	12.962	7.447	19.512
## 1605	2737	27.804	22.222	28.646	6.4236	12.386	6.536	16.146
## 1606	2767	23.491	22.013	24.645	2.6320	12.179	3.774	17.062
## 1607	1094	24.589	11.538	23.529	11.9910	17.367	13.462	19.118
## 1700	564	25.355	38.462	31.111	-7.3504	20.567	19.231	17.778
## 1701	49	14.286	NaN	33.333	NaN	18.367	NaN	33.333
## 1702	852	17.371	13.043	18.421	5.3776	17.371	10.870	18.421
## 1703	493	13.996	15.385	19.355	3.9702	12.170	7.692	20.968
## 1704	215	13.023	22.222	17.647	-4.5752	15.814	22.222	17.647
## 1705	814	19.165	3.704	26.087	22.3833	19.779	0.000	17.391
## 1706	1991	22.501	15.873	21.429	5.5556	19.990	9.524	25.893
## 1707	307	10.749	16.667	12.500	-4.1667	8.469	16.667	15.625
## 1708	380	16.053	20.000	22.727	2.7273	14.211	10.000	4.545
## 1709	376	36.436	25.000	33.333	8.3333	34.309	12.500	26.190
## 1710	941	26.142	15.625	35.526	19.9013	24.230	9.375	27.632
## 1711	413	6.780	8.696	7.317	-1.3786	7.506	8.696	4.878
## 1712	1338	14.723	7.812	13.333	5.5208	13.303	6.250	17.143
## 1800	281	25.623	0.000	35.714	35.7143	21.708	0.000	25.000
## 1801	7	0.000	NaN	0.000	NaN	14.286	NaN	0.000
## 1802	403	29.529	21.429	28.571	7.1429	28.536	21.429	25.000
## 1803	592	20.946	12.500	18.868	6.3679	17.905	12.500	13.208
## 1804	619	12.439	3.704	16.923	13.2194	13.247	3.704	13.846
## 1900	612	24.183	22.222	16.667	-5.5556	17.974	18.519	16.667
## 1901	630	16.825	2.703	18.421	15.7183	10.000	5.405	18.421
## 1902	982	24.134	11.628	21.277	9.6487	14.053	11.628	14.894
## 1903	181	13.260	0.000	18.182	18.1818	11.602	0.000	13.636
## 1904	1181	28.704	12.500	28.319	15.8186	23.793	18.750	35.398
## 1905	210	11.905	0.000	21.053	21.0526	10.952	9.091	10.526
## 1906	1439	17.929	6.452	20.000	13.5484	14.246	9.677	17.895
## 1907	1330	14.962	4.082	23.077	18.9953	13.684	4.082	20.513
## 1908	1192	15.185	4.348	18.182	13.8340	10.487	13.043	15.584
## 1909	587	10.562	18.182	4.082	-14.1002	5.622	0.000	12.245
## 1910	1713	26.853	18.333	30.935	12.6019	19.965	15.000	28.777
## 1911	850	27.412	18.750	25.000	6.2500	23.059	22.917	30.000
## 1912	800	17.375	11.905	26.154	14.2491	10.875	14.286	12.308
## 1913	145	21.379	14.286	38.462	24.1758	19.310	14.286	30.769

## 2000	597	21.943	16.667	21.739	5.0725	20.771	8.333	24.638
## 2001	406	33.005	35.714	28.571	-7.1429	28.571	28.571	26.531
## 2002	4181	20.139	11.189	26.923	15.7343	19.636	7.692	25.721
## 2003	1453	21.335	21.875	25.150	3.2747	19.683	12.500	19.760
## 2100	324	17.593	0.000	9.434	9.4340	15.741	0.000	9.434
## 2101	51	13.725	NaN	0.000	NaN	21.569	NaN	28.571
## 2102	442	13.801	10.000	9.804	-0.1961	8.371	5.000	7.843
## 2103	364	15.659	10.526	11.429	0.9023	9.890	5.263	11.429
## 2104	102	18.627	0.000	16.667	16.6667	10.784	0.000	33.333
## 2105	570	19.123	5.556	24.468	18.9125	13.684	0.000	17.021
## 2200	852	23.944	22.388	29.412	7.0237	17.723	4.478	19.118
## 2201	221	25.792	0.000	16.667	16.6667	16.742	0.000	5.556
## 2202	358	11.173	14.286	24.138	9.8522	8.939	7.143	6.897
## 2203	128	18.750	0.000	20.000	20.0000	9.375	0.000	13.333
## 2204	599	30.718	41.176	27.957	-13.2195	16.694	29.412	17.204
## 2205	919	17.954	20.000	20.732	0.7317	13.058	3.333	12.195
## 2206	115	6.957	25.000	0.000	-25.0000	8.696	25.000	0.000
## 2207	751	9.055	8.108	10.448	2.3397	6.924	8.108	13.433
## 2208	1999	8.454	10.976	9.091	-1.8847	6.753	7.317	6.818
## 2209	806	22.208	0.000	35.227	35.2273	15.012	33.333	19.318
## 2210	1243	14.240	13.462	13.605	0.1439	9.413	3.846	13.605
## 2211	964	16.805	14.286	16.364	2.0779	11.618	4.762	14.545
## 2212	247	13.360	14.286	6.897	-7.3892	12.955	14.286	3.448
## 2213	513	33.918	14.286	45.161	30.8756	26.706	28.571	29.032
## 2214	126	26.190	10.000	50.000	40.0000	21.429	20.000	50.000
## 2215	545	22.569	14.286	26.415	12.1294	15.963	0.000	18.868
## 2216	247	22.672	0.000	23.810	23.8095	17.814	0.000	14.286
## 2300	1906	30.745	25.000	32.895	7.8947	24.711	26.190	21.053
## 2301	385	41.299	28.571	40.000	11.4286	35.065	21.429	37.778
## 2302	396	23.737	6.250	26.667	20.4167	15.909	6.250	26.667
## 2303	4671	32.477	17.986	35.142	17.1559	22.672	17.986	29.481
## 2304	1208	30.629	19.048	30.252	11.2045	19.619	14.286	21.849
## 2305	674	25.964	13.333	27.273	13.9394	18.843	0.000	20.455
## 2306	454	29.515	7.143	28.571	21.4286	25.330	14.286	23.810
## 2307	561	38.503	20.000	54.000	34.0000	20.856	26.667	24.000
## 2308	3059	34.390	30.864	34.672	3.8073	29.356	32.099	28.467
## 2309	1803	32.557	15.789	38.202	22.4128	24.293	13.158	29.213
## 2310	904	31.084	11.111	29.268	18.1572	22.124	0.000	21.951
## 2311	505	31.683	7.143	30.645	23.5023	21.782	0.000	25.806
## 2312	1647	22.404	5.556	26.582	21.0267	16.211	1.389	21.519
## 2400	484	39.463	27.778	27.586	-0.1916	22.314	11.111	24.138
## 2401	2	50.000	NaN	100.000	NaN	0.000	NaN	0.000
## 2402	640	38.906	43.333	42.000	-1.3333	18.750	23.333	26.000
## 2403	2746	49.490	38.514	54.688	16.1740	22.323	16.216	31.250
## 2404	1545	43.495	49.398	47.312	-2.0858	18.770	18.072	21.505
## 2405	744	49.328	50.980	44.898	-6.0824	23.925	13.725	28.571
## 2406	795	44.277	37.500	43.103	5.6034	16.604	15.000	29.310
## 2500	1849	20.714	13.333	26.471	13.1373	15.197	3.333	16.667
## 2501	93	22.581	25.000	60.000	35.0000	13.978	25.000	20.000
## 2502	529	32.892	27.273	38.462	11.1888	15.690	18.182	12.308

## 2503	649	24.037	10.256	31.667	21.4103	13.251	7.692	25.000
## 2504	1278	18.153	10.000	19.858	9.8582	11.346	5.000	17.021
## 2505	2094	23.161	14.286	20.130	5.8442	13.515	6.250	17.532
## 2506	689	21.626	11.429	22.078	10.6494	11.756	8.571	14.286
## 2507	902	24.279	3.922	20.270	16.3487	12.971	1.961	13.514
## 2508	1039	23.195	13.208	25.253	12.0450	11.935	3.774	17.172
## 2600	844	17.654	19.565	18.667	-0.8986	17.180	17.391	24.000
## 2601	189	22.222	0.000	18.519	18.5185	16.402	0.000	14.815
## 2602	295	12.542	0.000	14.286	14.2857	11.186	0.000	17.143
## 2603	203	8.867	18.182	7.143	-11.0390	8.374	9.091	7.143
## 2604	1424	13.834	13.636	12.500	-1.1364	12.430	13.636	12.500
## 2605	341	16.716	18.182	24.444	6.2626	14.956	27.273	22.222
## 2606	106	11.321	NaN	18.750	NaN	8.491	NaN	12.500
## 2607	278	18.705	9.091	24.000	14.9091	24.460	18.182	24.000
## 2608	177	11.864	25.000	21.429	-3.5714	9.605	37.500	21.429
## 2609	89	8.989	0.000	14.286	14.2857	7.865	0.000	14.286
## 2610	337	6.528	16.667	0.000	-16.6667	6.825	16.667	7.407
## 2611	609	13.793	12.500	11.268	-1.2324	11.330	12.500	16.901
## 2612	134	14.179	0.000	14.286	14.2857	7.463	0.000	21.429
## 2613	1229	13.426	8.772	17.037	8.2651	12.286	5.263	16.296
## 2614	545	12.477	10.345	18.750	8.4052	15.229	13.793	18.750
## 2700	7457	44.710	34.969	47.176	12.2068	34.250	27.607	36.656
## 2701	2179	54.475	33.708	66.260	32.5523	38.320	24.719	50.000
## 2702	306	46.732	40.000	35.000	-5.0000	28.431	13.333	35.000
## 2703	801	36.954	26.667	48.837	22.1705	28.714	33.333	33.721
## 2704	342	43.275	33.333	46.667	13.3333	22.807	23.810	36.667
## 2705	2805	28.699	16.418	28.821	12.4031	14.581	5.970	13.974
## 2706	1052	33.270	30.769	41.053	10.2834	24.430	7.692	27.368
## 2707	468	46.368	75.000	57.692	-17.3077	41.239	0.000	42.308
## 2708	921	44.191	36.585	57.500	20.9146	30.727	21.951	37.500
## 2709	6	66.667	NaN	100.000	NaN	66.667	NaN	100.000
## 2710	168	63.690	0.000	66.667	66.6667	34.524	0.000	22.222
## 2711	993	34.038	0.000	42.105	42.1053	29.406	0.000	28.571
## 2712	1412	51.062	32.812	62.143	29.3304	23.584	10.938	39.286
## 2713	898	59.800	43.333	71.074	27.7410	38.753	23.333	50.413
## 2714	1415	55.830	100.000	56.774	-43.2258	44.735	60.000	38.710
## 2715	894	34.452	32.258	39.726	7.4680	17.002	25.806	12.329
## 2716	790	56.329	40.909	68.493	27.5841	31.013	9.091	39.726
## 2717	1124	59.164	42.308	62.609	20.3010	45.196	57.692	43.478
## 2718	466	50.644	33.333	65.753	32.4201	35.408	33.333	39.726
## 2719	2269	58.131	55.000	63.670	8.6704	48.788	50.000	47.191
## 2720	844	43.365	28.125	51.111	22.9861	25.000	25.000	31.111
## 2721	522	35.441	25.806	38.235	12.4288	17.625	19.355	14.706
## 2722	355	43.944	58.333	37.500	-20.8333	30.423	25.000	28.125
## 2723	1305	49.808	31.481	56.818	25.3367	24.828	14.815	34.091
## 2724	795	48.553	26.471	54.930	28.4590	26.164	14.706	42.254
## 2725	1988	47.535	54.902	54.286	-0.6162	25.453	7.843	33.333
## 2726	776	47.938	50.000	57.143	7.1429	27.964	15.000	35.065
## 2727	732	43.033	37.500	63.636	26.1364	24.863	18.750	34.848
## 2728	3210	48.037	30.252	54.459	24.2065	30.685	25.210	35.669

## 2729	2102	60.324	34.737	72.330	37.5933	41.722	31.579	59.223
## 2730	2202	51.181	35.088	53.819	18.7317	30.200	24.561	31.597
## 2731	1898	37.039	37.500	35.200	-2.3000	24.025	22.222	27.200
## 2732	2816	32.528	22.059	38.650	16.5915	22.124	13.235	21.472
## 2733	963	46.833	27.660	49.206	21.5468	31.983	23.404	30.159
## 2734	1653	41.016	37.864	43.655	5.7907	28.433	30.097	26.904
## 2735	2639	56.196	38.806	60.912	22.1061	41.834	26.866	48.534
## 2736	1600	46.375	28.261	48.810	20.5487	33.875	21.739	31.548
## 2737	1891	36.700	27.835	38.144	10.3093	18.826	16.495	22.165
## 2738	5668	49.665	33.493	57.860	24.3667	39.591	37.799	38.462
## 2739	6220	56.849	52.198	63.988	11.7898	44.357	37.363	46.522
## 2740	1568	35.077	28.125	41.096	12.9709	19.324	15.625	23.288
## 2741	1641	27.179	16.393	32.298	15.9047	18.342	19.672	19.255
## 2742	1429	67.460	52.632	69.504	16.8720	52.974	47.368	49.645
## 2743	546	60.073	44.444	61.702	17.2577	39.744	22.222	55.319
## 2744	4	0.000	NaN	NaN	NaN	0.000	NaN	NaN
## 2745	349	49.284	12.500	50.000	37.5000	30.659	12.500	30.556
## 2746	3202	21.830	10.084	31.985	21.9013	12.586	9.244	18.015
## 2747	550	43.273	21.429	56.000	34.5714	24.364	21.429	42.000
## 2748	390	36.667	8.333	46.667	38.3333	28.974	8.333	24.444
## 2800	1830	43.115	38.554	46.667	8.1124	23.552	21.687	27.407
## 2801	133	54.887	NaN	61.111	NaN	41.353	NaN	38.889
## 2802	776	51.804	32.143	58.537	26.3937	34.665	50.000	34.146
## 2803	504	47.817	26.667	51.020	24.3537	26.190	26.667	32.653
## 2804	963	47.352	48.718	40.698	-8.0203	23.053	28.205	26.744
## 2805	709	56.559	51.724	54.737	3.0127	36.530	37.931	36.842
## 2806	322	61.180	66.667	59.459	-7.2072	47.826	50.000	54.054
## 2807	124	41.935	33.333	47.368	14.0351	32.258	50.000	15.789
## 2808	1723	46.257	36.735	46.471	9.7359	28.265	24.490	24.706
## 2809	951	37.855	38.235	36.047	-2.1888	24.606	29.412	30.233
## 2900	2558	77.170	65.517	77.500	11.9828	71.970	67.241	75.000
## 2901	37	70.270	NaN	50.000	NaN	64.865	NaN	50.000
## 2902	612	69.608	18.182	79.167	60.9848	52.614	9.091	60.417
## 2903	13	53.846	NaN	NaN	NaN	46.154	NaN	NaN
## 2904	18	72.222	NaN	100.000	NaN	61.111	NaN	0.000
## 2905	573	67.016	NaN	70.492	NaN	56.894	NaN	65.574
## 2906	275	81.455	80.000	82.353	2.3529	71.273	60.000	58.824
## 2907	486	57.819	50.000	70.213	20.2128	52.263	33.333	51.064
## 2908	28	71.429	NaN	100.000	NaN	71.429	NaN	80.000
## 2909	221	57.014	33.333	69.444	36.1111	50.679	33.333	55.556
## 2910	310	48.710	42.857	63.830	20.9726	48.065	42.857	53.191
## 2911	341	58.944	41.176	76.471	35.2941	57.185	47.059	64.706
## 2912	278	92.086	66.667	93.750	27.0833	74.460	66.667	68.750
## 2913	254	95.669	NaN	87.500	NaN	85.433	NaN	77.500
## 2914	123	79.675	33.333	77.273	43.9394	73.171	33.333	72.727
## 2915	1	100.000	NaN	NaN	NaN	100.000	NaN	NaN
## 2916	1264	64.794	41.176	77.586	36.4097	42.326	23.529	56.322
## 2917	80	86.250	100.000	71.429	-28.5714	76.250	100.000	78.571
## 2918	0	NA	NA	NA	NA	NA	NA	NA
## 2919	186	80.108	66.667	75.000	8.3333	74.194	100.000	70.833

## 2920	49	44.898	NaN	50.000	NaN	48.980	NaN	75.000
## 2921	429	64.569	0.000	53.846	53.8462	58.275	0.000	59.615
## 2922	133	86.466	NaN	100.000	NaN	81.955	NaN	100.000
## 2923	57	61.404	NaN	60.000	NaN	56.140	NaN	40.000
## 3000	217	52.995	33.333	12.500	-20.8333	36.866	50.000	25.000
## 3001	18	44.444	NaN	100.000	NaN	44.444	NaN	60.000
## 3002	1119	31.099	32.653	28.125	-4.5281	15.550	8.163	15.625
## 3003	852	40.141	50.000	42.647	-7.3529	26.526	19.231	26.471
## 3004	2441	42.155	33.696	46.386	12.6899	24.211	20.109	22.892
## 3005	823	41.069	10.714	53.448	42.7340	26.002	39.286	29.310
## 3100	1396	11.533	0.000	9.524	9.5238	8.095	0.000	7.483
## 3101	299	12.709	0.000	5.000	5.0000	12.040	20.000	20.000
## 3102	338	19.231	14.286	17.391	3.1056	13.905	0.000	13.043
## 3103	477	16.143	11.111	28.261	17.1498	12.159	14.815	10.870
## 3104	2718	16.924	10.476	18.727	8.2504	10.927	7.619	13.483
## 3105	562	15.125	16.667	11.111	-5.5556	14.413	11.111	17.778
## 3106	393	11.959	0.000	10.000	10.0000	7.634	0.000	5.000
## 3107	1257	10.979	4.762	10.219	5.4571	12.172	1.587	8.759
## 3108	246	15.041	14.286	8.696	-5.5901	15.041	0.000	13.043
## 3109	416	7.452	13.636	0.000	-13.6364	6.731	13.636	12.121
## 3110	509	29.470	13.043	28.889	15.8454	12.967	0.000	20.000
## 3200	2287	47.005	42.574	54.774	12.1996	41.758	45.545	43.719
## 3201	550	49.818	47.619	58.824	11.2045	44.909	38.095	47.059
## 3202	1305	53.640	65.000	60.177	-4.8230	45.134	45.000	46.903
## 3203	2247	58.923	37.333	63.071	25.7372	47.085	42.667	49.793
## 3204	2761	65.339	49.091	72.852	23.7613	55.668	50.000	64.948
## 3205	1835	50.518	36.585	58.683	22.0973	39.455	40.244	37.725
## 3206	746	54.826	58.824	59.420	0.5968	43.298	70.588	49.275
## 3207	1795	55.543	53.704	63.187	9.4831	50.696	55.556	47.253
## 3300	1251	42.446	20.690	44.286	23.5961	40.528	17.241	40.714
## 3301	1482	53.981	28.814	60.784	31.9708	51.754	28.814	60.784
## 3302	583	29.503	0.000	39.286	39.2857	27.273	0.000	46.429
## 3303	2004	35.030	25.000	44.211	19.2105	33.283	26.471	41.053
## 3304	9371	56.205	40.072	64.280	24.2077	52.609	36.101	59.398
## 3305	5048	35.796	25.143	41.365	16.2217	33.796	24.571	39.019
## 3306	2238	60.411	42.857	64.259	21.4014	54.021	39.286	55.894
## 3307	528	38.258	23.810	66.038	42.2282	32.955	23.810	32.075
## 3308	3329	44.338	44.660	49.885	5.2249	42.295	47.573	45.057
## 3309	1495	51.371	28.261	53.719	25.4581	49.833	26.087	57.851
## 3310	2380	55.840	55.319	58.635	3.3154	51.471	56.383	51.406
## 3311	473	30.867	25.000	46.296	21.2963	28.330	50.000	37.037
## 3312	6487	43.333	30.645	44.334	13.6891	41.807	32.796	42.918
## 3313	554	29.061	8.333	30.612	22.2789	25.090	8.333	24.490
## 3314	1520	46.908	33.962	55.814	21.8517	46.053	32.075	55.233
## 3315	1340	48.358	48.000	54.667	6.6667	47.985	52.000	51.111
## 3316	2919	45.632	27.586	48.966	21.3793	44.947	32.759	48.736
## 3317	459	45.316	40.909	55.556	14.6465	46.187	40.909	61.111
## 3318	757	80.449	90.323	81.944	-8.3781	77.015	90.323	77.778
## 3319	501	61.876	44.444	74.026	29.5815	59.481	44.444	63.636
## 3320	1967	31.469	21.875	31.788	9.9131	30.910	21.875	31.788

## 3321	815	36.933	20.833	42.857	22.0238	34.847	29.167	37.662
## 3322	604	34.934	14.286	40.594	26.3083	32.616	14.286	31.683
## 3400	785	47.898	33.333	58.025	24.6914	25.096	23.333	28.395
## 3401	22	77.273	0.000	100.000	100.0000	40.909	0.000	33.333
## 3402	22	63.636	100.000	100.000	0.0000	31.818	100.000	100.000
## 3403	118	48.305	25.000	47.368	22.3684	19.492	25.000	31.579
## 3404	119	49.580	20.000	80.000	60.0000	34.454	20.000	80.000
## 3500	340	28.529	20.000	46.667	26.6667	17.647	20.000	36.667
## 3501	5	80.000	NaN	100.000	NaN	20.000	NaN	0.000
## 3502	0	NA	NA	NA	NA	NA	NA	NA
## 3503	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
## 3504	100	22.000	50.000	20.000	-30.0000	10.000	16.667	6.667
## 3505	19	36.842	NaN	0.000	NaN	10.526	NaN	0.000
## 3506	22	50.000	100.000	NaN	NaN	9.091	0.000	NaN
## 3600	299	60.870	25.000	66.667	41.6667	56.522	50.000	44.444
## 3601	294	42.177	NaN	65.517	NaN	36.054	NaN	48.276
## 3602	79	32.911	NaN	10.000	NaN	27.848	NaN	10.000
## 3603	95	66.316	100.000	100.000	0.0000	54.737	66.667	62.500
## 3604	239	37.657	NaN	68.421	NaN	31.799	NaN	36.842
## 3605	48	45.833	NaN	42.857	NaN	33.333	NaN	28.571
## 3606	6	83.333	NaN	NaN	NaN	83.333	NaN	NaN
## 3607	128	34.375	100.000	50.000	-50.0000	33.594	0.000	50.000
## 3608	6	83.333	NaN	NaN	NaN	83.333	NaN	NaN
## 3609	430	84.186	83.333	83.721	0.3876	81.395	100.000	74.419
## 3610	408	40.196	26.316	52.632	26.3158	33.088	15.789	52.632
## 3611	332	59.036	100.000	51.163	-48.8372	53.313	100.000	48.837
## 3612	2591	47.974	40.506	50.530	10.0237	37.437	39.241	34.982
## 3613	91	28.571	0.000	0.000	0.0000	35.165	0.000	0.000
## 3614	334	26.347	14.286	31.818	17.5325	16.467	28.571	22.727
## 3615	0	NA	NA	NA	NA	NA	NA	NA
## 3616	844	78.673	57.143	77.922	20.7792	60.545	57.143	62.338
##	ChgN	1CtyMF96	1CtyMF14					
## 1000	16.5620	26	49					
## 1100	8.0338	86	473					
## 1101	3.2200	13	43					
## 1102	6.1530	68	163					
## 1103	5.3396	142	292					
## 1104	9.3591	142	338					
## 1105	3.4899	275	817					
## 1106	6.9729	57	118					
## 1107	10.2273	24	88					
## 1108	0.8065	8	31					
## 1109	-7.2727	44	80					
## 1110	-3.5254	132	254					
## 1111	8.9260	37	76					
## 1200	37.7143	7	50					
## 1201	10.6129	138	288					
## 1202	11.7408	80	382					
## 1203	-7.8392	84	212					
## 1204	12.2807	6	57					

## 1205	-13.1579	2	19
## 1206	NaN	0	7
## 1207	19.4035	28	97
## 1208	-3.3333	15	150
## 1209	NaN	0	22
## 1210	24.1525	12	59
## 1211	3.4388	53	186
## 1212	16.4835	14	91
## 1213	45.8065	4	155
## 1300	4.3712	71	502
## 1301	NaN	0	2
## 1302	-25.0000	14	28
## 1303	10.2188	353	442
## 1304	5.2377	86	113
## 1305	11.3428	49	138
## 1306	8.9829	57	158
## 1307	2.6114	217	231
## 1308	-6.7982	76	96
## 1309	15.2683	50	82
## 1310	12.9110	95	113
## 1311	9.8729	181	300
## 1312	6.1863	259	356
## 1313	2.9762	24	84
## 1314	2.4032	168	241
## 1315	11.0294	24	34
## 1400	25.5528	37	110
## 1401	-7.9545	16	44
## 1402	17.8571	35	112
## 1403	21.5494	60	199
## 1404	16.6667	15	30
## 1405	18.6720	29	94
## 1406	38.5714	21	150
## 1407	14.8099	67	168
## 1408	21.3896	47	226
## 1409	29.5500	31	162
## 1410	33.3333	24	48
## 1500	14.0186	70	123
## 1501	0.0000	3	3
## 1502	-5.1030	24	89
## 1503	13.6905	48	56
## 1504	-100.0000	1	2
## 1505	8.0000	26	25
## 1506	36.3636	1	22
## 1507	7.6923	5	13
## 1508	33.3333	2	6
## 1600	7.7695	290	363
## 1601	40.0000	1	5
## 1602	7.9129	71	110
## 1603	17.3077	23	52
## 1604	12.0654	94	82

## 1605	9.6099	153	192
## 1606	13.2880	159	211
## 1607	5.6561	52	68
## 1700	-1.4530	26	45
## 1701	NaN	0	3
## 1702	7.5515	46	76
## 1703	13.2754	13	62
## 1704	-4.5752	9	17
## 1705	17.3913	27	69
## 1706	16.3690	63	224
## 1707	-1.0417	6	32
## 1708	-5.4545	20	22
## 1709	13.6905	8	42
## 1710	18.2566	32	76
## 1711	-3.8176	23	41
## 1712	10.8929	64	105
## 1800	25.0000	4	28
## 1801	NaN	0	1
## 1802	3.5714	14	28
## 1803	0.7075	24	53
## 1804	10.1425	27	65
## 1900	-1.8519	27	48
## 1901	13.0156	37	38
## 1902	3.2657	43	94
## 1903	13.6364	5	22
## 1904	16.6482	32	113
## 1905	1.4354	11	19
## 1906	8.2173	62	95
## 1907	16.4312	49	78
## 1908	2.5409	46	77
## 1909	12.2449	11	49
## 1910	13.7770	60	139
## 1911	7.0833	48	40
## 1912	-1.9780	42	65
## 1913	16.4835	7	13
## 2000	16.3043	12	69
## 2001	-2.0408	14	49
## 2002	18.0288	143	416
## 2003	7.2605	32	167
## 2100	9.4340	3	53
## 2101	NaN	0	7
## 2102	2.8431	20	51
## 2103	6.1654	19	35
## 2104	33.3333	10	6
## 2105	17.0213	18	94
## 2200	14.6400	67	68
## 2201	5.5556	4	18
## 2202	-0.2463	14	29
## 2203	13.3333	1	15
## 2204	-12.2075	17	93

## 2205	8.8618	30	82
## 2206	-25.0000	4	9
## 2207	5.3247	37	67
## 2208	-0.4989	82	176
## 2209	-14.0152	3	88
## 2210	9.7593	52	147
## 2211	9.7835	21	110
## 2212	-10.8374	7	29
## 2213	0.4608	14	62
## 2214	30.0000	10	6
## 2215	18.8679	14	53
## 2216	14.2857	10	21
## 2300	-5.1378	84	152
## 2301	16.3492	14	45
## 2302	20.4167	16	45
## 2303	11.4955	139	424
## 2304	7.5630	42	119
## 2305	20.4545	15	88
## 2306	9.5238	14	63
## 2307	-2.6667	15	50
## 2308	-3.6316	81	274
## 2309	16.0556	38	178
## 2310	21.9512	18	123
## 2311	25.8065	14	62
## 2312	20.1301	72	158
## 2400	13.0268	18	58
## 2401	NaN	0	1
## 2402	2.6667	30	50
## 2403	15.0338	148	192
## 2404	3.4331	83	93
## 2405	14.8459	51	49
## 2406	14.3103	40	58
## 2500	13.3333	60	204
## 2501	-5.0000	4	5
## 2502	-5.8741	11	65
## 2503	17.3077	39	60
## 2504	12.0213	40	141
## 2505	11.2825	112	154
## 2506	5.7143	35	77
## 2507	11.5527	51	74
## 2508	13.3981	53	99
## 2600	6.6087	46	75
## 2601	14.8148	8	27
## 2602	17.1429	7	35
## 2603	-1.9481	11	28
## 2604	-1.1364	66	144
## 2605	-5.0505	11	45
## 2606	NaN	0	16
## 2607	5.8182	11	25
## 2608	-16.0714	8	14

## 2609	14.2857	3	7
## 2610	-9.2593	18	27
## 2611	4.4014	24	71
## 2612	21.4286	1	14
## 2613	11.0331	57	135
## 2614	4.9569	29	32
## 2700	9.0482	163	903
## 2701	25.2809	89	246
## 2702	21.6667	15	20
## 2703	0.3876	15	86
## 2704	12.8571	21	30
## 2705	8.0036	67	229
## 2706	19.6761	13	95
## 2707	42.3077	4	52
## 2708	15.5488	41	80
## 2709	NaN	0	2
## 2710	22.2222	6	9
## 2711	28.5714	3	133
## 2712	28.3482	64	140
## 2713	27.0799	30	121
## 2714	-21.2903	5	155
## 2715	-13.4777	31	73
## 2716	30.6351	22	73
## 2717	-14.2140	26	115
## 2718	6.3927	3	73
## 2719	-2.8090	20	267
## 2720	6.1111	32	90
## 2721	-4.6490	31	34
## 2722	3.1250	12	32
## 2723	19.2761	54	132
## 2724	27.5476	34	71
## 2725	25.4902	51	210
## 2726	20.0649	20	77
## 2727	16.0985	32	66
## 2728	10.4587	119	314
## 2729	27.6444	95	206
## 2730	7.0358	57	288
## 2731	4.9778	72	125
## 2732	8.2371	68	326
## 2733	6.7545	47	63
## 2734	-3.1935	103	197
## 2735	21.6685	67	307
## 2736	9.8085	46	168
## 2737	5.6701	97	194
## 2738	0.6625	209	598
## 2739	9.1598	182	647
## 2740	7.6627	32	146
## 2741	-0.4175	61	161
## 2742	2.2770	38	141
## 2743	33.0969	18	47

## 2744	NaN	0	0
## 2745	18.0556	8	36
## 2746	8.7710	119	272
## 2747	20.5714	14	50
## 2748	16.1111	12	45
## 2800	5.7207	83	135
## 2801	NaN	0	18
## 2802	-15.8537	28	82
## 2803	5.9864	15	49
## 2804	-1.4609	39	86
## 2805	-1.0889	29	95
## 2806	4.0541	6	37
## 2807	-34.2105	6	19
## 2808	0.2161	49	170
## 2809	0.8208	34	86
## 2900	7.7586	58	200
## 2901	NaN	0	4
## 2902	51.3258	11	48
## 2903	NaN	0	0
## 2904	NaN	0	1
## 2905	NaN	0	61
## 2906	-1.1765	5	17
## 2907	17.7305	6	47
## 2908	NaN	0	5
## 2909	22.2222	3	36
## 2910	10.3343	7	47
## 2911	17.6471	17	17
## 2912	2.0833	6	16
## 2913	NaN	0	40
## 2914	39.3939	3	22
## 2915	NaN	0	0
## 2916	32.7924	34	174
## 2917	-21.4286	1	14
## 2918	NA	NA	NA
## 2919	-29.1667	3	24
## 2920	NaN	0	4
## 2921	59.6154	1	52
## 2922	NaN	0	5
## 2923	NaN	0	5
## 3000	-25.0000	6	8
## 3001	NaN	0	5
## 3002	7.4617	49	96
## 3003	7.2398	52	68
## 3004	2.7829	184	166
## 3005	-9.9754	28	58
## 3100	7.4830	8	147
## 3101	0.0000	5	20
## 3102	13.0435	14	23
## 3103	-3.9452	27	46
## 3104	5.8641	105	267

## 3105	6.6667	18	45
## 3106	5.0000	4	40
## 3107	7.1718	63	137
## 3108	13.0435	7	23
## 3109	-1.5152	22	33
## 3110	20.0000	23	45
## 3200	-1.8260	101	199
## 3201	8.9636	21	51
## 3202	1.9027	40	113
## 3203	7.1259	75	241
## 3204	14.9485	110	291
## 3205	-2.5194	82	167
## 3206	-21.3129	17	69
## 3207	-8.3028	54	182
## 3300	23.4729	29	140
## 3301	31.9708	59	204
## 3302	46.4286	9	56
## 3303	14.5820	68	190
## 3304	23.2968	277	1229
## 3305	14.4478	175	469
## 3306	16.6078	56	263
## 3307	8.2659	21	53
## 3308	-2.5153	103	435
## 3309	31.7643	46	121
## 3310	-4.9774	94	249
## 3311	-12.9630	4	54
## 3312	10.1221	186	706
## 3313	16.1565	12	49
## 3314	23.1571	53	172
## 3315	-0.8889	25	225
## 3316	15.9770	58	435
## 3317	20.2020	22	36
## 3318	-12.5448	31	72
## 3319	19.1919	18	77
## 3320	9.9131	32	302
## 3321	8.4957	24	77
## 3322	17.3975	7	101
## 3400	5.0617	30	81
## 3401	33.3333	1	3
## 3402	0.0000	2	1
## 3403	6.5789	4	19
## 3404	60.0000	5	5
## 3500	16.6667	30	30
## 3501	NaN	0	1
## 3502	NA	NA	NA
## 3503	NaN	0	0
## 3504	-10.0000	6	15
## 3505	NaN	0	2
## 3506	NaN	1	0
## 3600	-5.5556	8	9


```

## 3601      NaN      0      29
## 3602      NaN      0      10
## 3603    -4.1667      3       8
## 3604      NaN      0      19
## 3605      NaN      0       7
## 3606      NaN      0       0
## 3607    50.0000      2      18
## 3608      NaN      0       0
## 3609   -25.5814      6      43
## 3610    36.8421     19      19
## 3611   -51.1628      1      43
## 3612    -4.2582     79     283
## 3613     0.0000      1       2
## 3614    -5.8442      7      22
## 3615        NA      NA      NA
## 3616     5.1948     21      77

```

```
print("GenderTeamSize in 2018")
```

```
## [1] "GenderTeamSize in 2018"
```

```
print(GenderTeamSize)
```

```

##      Articles FirstF FirstM   FirstP LastF LastM   LastP
## 1000      259  4.365  3.702 8.595e-02 4.139 3.941 4.676e-01
## 1100      404  4.367  3.395 9.340e-05 4.181 3.693 4.666e-02
## 1101       39  2.647  2.253 4.566e-01 2.069 2.503 4.445e-01
## 1102      181  4.159  3.461 1.136e-01 4.542 3.511 6.200e-02
## 1103      313  3.583  3.117 6.276e-02 3.062 3.504 1.002e-01
## 1104      350  3.835  3.389 3.915e-02 3.765 3.512 2.016e-01
## 1105      764  3.606  3.211 2.401e-02 3.430 3.359 7.500e-01
## 1106      232  4.724  3.791 9.578e-03 4.509 4.244 4.081e-01
## 1107       88  3.521  3.412 9.860e-01 3.791 3.371 3.717e-01
## 1108       30  3.722  3.590 8.072e-01 4.478 3.438 1.954e-01
## 1109       76  3.817  2.606 1.531e-02 3.753 2.995 4.288e-01
## 1110      233  3.533  3.431 9.942e-01 3.577 3.433 6.261e-01
## 1111      135  3.611  3.391 8.287e-01 3.953 3.352 2.407e-01
## 1200       34  1.260  1.441 5.159e-01 1.557 1.245 2.262e-01
## 1201      387  2.096  1.884 1.192e-01 2.002 2.013 7.993e-01
## 1202      410  1.268  1.143 1.706e-02 1.189 1.217 3.300e-01
## 1203      210  2.239  1.556 3.745e-04 2.084 1.798 1.690e-01
## 1204       72  2.255  2.155 7.887e-01 2.029 2.387 3.542e-01
## 1205        2  1.000  1.000      NaN 1.000 1.000      NaN
## 1206       26  1.260  1.571 3.077e-01 1.356 1.422 6.965e-01
## 1207      100  1.616  1.218 6.827e-03 1.455 1.376 5.651e-01
## 1208      186  1.141  1.167 7.147e-01 1.160 1.140 8.216e-01
## 1209       22  1.173  1.537 2.219e-01 1.281 1.364 7.080e-01
## 1210       45  1.514  1.704 5.753e-01 1.376 1.834 7.337e-02
## 1211      210  1.388  1.182 2.337e-02 1.474 1.136 2.148e-04
## 1212      106  1.559  1.187 1.331e-02 1.600 1.150 9.898e-04
## 1213      182  1.311  1.380 5.558e-01 1.300 1.395 1.644e-01

```

## 1300	470	4.966	3.896	9.779e-05	4.673	4.296	2.218e-01
## 1301	5	4.472	5.192	7.609e-01	5.000	4.865	1.000e+00
## 1302	30	3.736	6.124	8.161e-03	4.424	5.382	3.097e-01
## 1303	283	5.323	4.650	3.488e-02	5.058	4.870	4.868e-01
## 1304	115	3.931	3.682	5.193e-01	3.711	3.807	8.794e-01
## 1305	132	4.499	4.342	4.921e-01	4.409	4.395	8.889e-01
## 1306	142	6.947	6.130	3.843e-01	6.303	6.764	6.903e-01
## 1307	185	5.819	5.410	3.946e-01	5.134	5.798	1.963e-01
## 1308	62	4.868	5.198	7.556e-01	5.357	4.956	2.617e-01
## 1309	71	4.717	4.524	5.139e-01	4.480	4.761	6.427e-01
## 1310	110	4.922	5.515	2.232e-01	4.962	5.275	5.080e-01
## 1311	322	4.886	4.431	3.108e-01	5.143	4.443	8.440e-02
## 1312	326	5.268	4.083	2.155e-03	4.959	4.355	2.002e-01
## 1313	127	5.816	5.743	9.610e-01	4.669	6.261	1.428e-02
## 1314	232	4.369	3.998	2.578e-01	4.285	4.095	3.529e-01
## 1315	35	6.030	3.829	5.194e-02	4.726	4.630	1.000e+00
## 1400	100	2.338	1.607	4.252e-04	2.210	1.715	1.694e-02
## 1401	56	2.642	2.108	2.352e-01	2.449	2.251	7.287e-01
## 1402	93	2.199	1.830	1.357e-01	2.074	1.898	4.867e-01
## 1403	181	2.106	1.918	3.335e-01	2.105	1.915	2.824e-01
## 1404	37	1.888	1.898	1.000e+00	2.176	1.790	4.234e-01
## 1405	81	2.092	1.825	2.746e-01	1.960	1.921	9.274e-01
## 1406	129	2.125	2.107	8.227e-01	2.051	2.182	5.395e-01
## 1407	121	1.870	1.754	4.377e-01	1.836	1.781	6.947e-01
## 1408	212	2.070	2.347	9.244e-02	2.208	2.259	8.860e-01
## 1409	151	1.785	2.107	5.612e-02	1.813	2.061	1.123e-01
## 1410	43	1.794	1.528	3.801e-01	1.595	1.634	9.547e-01
## 1500	145	4.022	3.465	1.994e-01	3.774	3.577	3.359e-01
## 1501	15	3.464	3.576	9.460e-01	3.464	3.559	1.000e+00
## 1502	66	5.599	4.684	1.609e-01	3.992	5.158	5.843e-02
## 1503	62	4.386	4.060	3.993e-01	3.596	4.320	3.717e-01
## 1504	5	5.313	4.472	3.329e-01	5.000	4.949	1.000e+00
## 1505	18	4.478	4.140	6.475e-01	3.107	4.501	1.998e-01
## 1506	14	4.738	3.726	4.154e-01	5.635	3.477	1.033e-01
## 1507	26	3.868	2.574	2.059e-01	3.000	2.868	8.921e-01
## 1508	31	3.923	3.464	3.203e-01	3.706	3.582	9.583e-01
## 1600	355	4.551	3.719	1.469e-02	3.959	3.940	9.334e-01
## 1601	22	3.766	4.914	1.761e-01	5.087	4.247	5.741e-01
## 1602	87	4.789	4.047	1.859e-01	4.424	4.320	9.555e-01
## 1603	25	4.924	4.529	7.298e-01	4.036	4.887	4.186e-01
## 1604	61	4.242	3.746	5.254e-01	4.307	3.768	4.910e-01
## 1605	138	5.109	4.515	2.033e-01	4.630	4.725	7.997e-01
## 1606	158	4.005	3.332	1.459e-01	3.759	3.476	6.734e-01
## 1607	52	4.366	4.821	5.160e-01	4.638	4.618	7.883e-01
## 1700	54	2.451	2.180	4.881e-01	2.166	2.293	8.180e-01
## 1701	10	3.873	2.611	3.426e-01	3.464	2.685	5.876e-01
## 1702	9	1.414	2.737	1.746e-01	2.520	2.289	5.946e-01
## 1703	36	3.870	2.216	5.867e-02	2.621	2.395	6.326e-01
## 1704	23	2.034	2.885	4.952e-01	2.146	2.870	5.313e-01
## 1705	41	3.174	2.254	9.231e-02	2.874	2.311	4.410e-01

## 1706	98	3.185	2.921	6.603e-01	2.793	3.038	6.042e-01
## 1707	19	3.224	2.721	6.783e-01	3.464	2.752	4.486e-01
## 1708	26	2.289	2.521	6.690e-01	2.639	2.459	8.083e-01
## 1709	63	2.136	2.693	1.299e-01	2.353	2.563	5.916e-01
## 1710	84	2.517	2.568	9.798e-01	2.314	2.627	4.611e-01
## 1711	37	4.559	2.374	1.929e-02	4.899	2.354	5.587e-02
## 1712	96	3.298	2.995	5.307e-01	3.587	2.893	2.635e-02
## 1800	31	3.112	2.727	7.142e-01	3.080	2.772	6.239e-01
## 1801	NA	NA	NA	NA	NA	NA	NA
## 1802	43	2.847	2.636	7.976e-01	2.996	2.584	6.157e-01
## 1803	54	2.155	2.722	1.053e-01	2.104	2.652	5.080e-01
## 1804	43	1.888	1.826	8.259e-01	2.460	1.763	1.804e-01
## 1900	66	3.198	2.871	5.774e-01	3.220	2.871	5.070e-01
## 1901	53	3.796	3.049	2.316e-01	2.737	3.377	2.083e-01
## 1902	110	2.814	2.840	7.957e-01	2.376	3.027	1.684e-01
## 1903	25	1.644	3.259	1.420e-02	1.681	3.211	3.049e-02
## 1904	154	2.178	2.400	3.540e-01	1.668	2.735	8.776e-06
## 1905	26	3.012	3.550	2.972e-01	3.293	3.347	7.250e-01
## 1906	103	3.853	3.035	2.171e-02	3.530	3.212	5.209e-01
## 1907	118	3.816	3.657	8.159e-01	3.561	3.761	7.722e-01
## 1908	56	3.354	2.878	4.259e-01	2.667	3.069	4.124e-01
## 1909	134	3.377	2.668	7.889e-02	3.185	2.705	2.140e-01
## 1910	169	3.824	3.125	3.821e-02	3.776	3.328	2.023e-01
## 1911	51	3.138	1.943	1.778e-02	1.698	2.666	1.329e-01
## 1912	62	2.854	2.713	8.669e-01	1.861	2.821	2.732e-01
## 1913	17	4.025	4.117	7.880e-01	3.201	4.529	3.607e-01
## 2000	65	1.560	1.563	9.858e-01	1.642	1.535	6.210e-01
## 2001	67	2.487	2.125	1.356e-01	2.304	2.210	6.833e-01
## 2002	381	1.869	1.664	4.440e-02	1.759	1.708	6.359e-01
## 2003	140	2.054	1.799	1.835e-01	1.991	1.827	3.577e-01
## 2100	56	3.593	2.854	1.842e-01	2.910	3.055	6.380e-01
## 2101	16	1.000	2.593	8.522e-02	1.442	2.564	2.113e-01
## 2102	85	2.030	2.852	4.737e-02	2.038	2.823	4.621e-02
## 2103	58	2.161	3.264	3.921e-02	2.586	3.141	2.574e-01
## 2104	31	2.234	2.095	5.484e-01	2.259	2.090	7.587e-01
## 2105	182	2.930	3.169	4.529e-01	2.799	3.189	3.147e-01
## 2200	42	2.782	2.805	9.053e-01	2.539	2.860	7.788e-01
## 2201	36	2.395	2.619	6.135e-01	2.284	2.656	4.128e-01
## 2202	36	1.861	2.573	2.159e-01	1.000	2.547	1.674e-01
## 2203	30	2.948	2.229	2.828e-01	3.009	2.328	3.689e-01
## 2204	133	5.058	4.516	3.537e-01	4.417	4.737	2.797e-01
## 2205	221	3.231	2.557	6.217e-03	3.023	2.630	1.642e-01
## 2206	12	3.000	2.989	1.000e+00	NA	NA	NA
## 2207	45	3.302	2.795	6.962e-01	2.000	2.872	8.134e-02
## 2208	99	2.867	2.903	9.319e-01	2.985	2.891	8.763e-01
## 2209	129	3.593	2.932	6.123e-02	2.986	3.087	7.694e-01
## 2210	193	4.083	2.798	6.039e-04	3.691	2.883	7.941e-02
## 2211	126	4.306	2.975	5.658e-03	4.051	3.083	1.605e-01
## 2212	49	3.970	2.662	7.326e-02	5.144	2.792	2.314e-02
## 2213	89	3.322	2.970	3.903e-01	2.942	3.150	4.625e-01

## 2214	14	1.587	1.646	9.311e-01	2.000	1.506	5.826e-01
## 2215	152	3.339	2.579	3.181e-03	2.766	2.715	9.788e-01
## 2216	44	1.707	2.045	3.455e-01	1.381	2.164	2.393e-02
## 2300	195	3.089	2.648	3.795e-02	2.654	2.845	4.718e-01
## 2301	95	2.293	2.536	4.629e-01	2.090	2.707	4.917e-02
## 2302	48	3.256	4.312	1.815e-01	3.457	4.228	3.570e-01
## 2303	487	3.548	3.494	9.863e-01	3.426	3.552	8.292e-01
## 2304	179	4.387	3.884	1.192e-01	4.085	4.063	7.061e-01
## 2305	155	4.183	3.490	5.614e-02	4.159	3.557	2.484e-01
## 2306	95	3.382	3.203	7.814e-01	2.962	3.408	4.054e-01
## 2307	161	4.451	3.901	2.030e-01	4.171	4.269	9.150e-01
## 2308	379	2.526	2.440	5.278e-01	2.198	2.640	2.049e-02
## 2309	202	3.198	3.571	1.306e-01	3.007	3.574	1.281e-01
## 2310	176	4.510	3.732	1.294e-02	4.054	3.994	8.963e-01
## 2311	117	3.802	4.026	8.546e-01	3.985	3.925	7.380e-01
## 2312	189	3.481	3.104	1.939e-01	3.248	3.222	7.842e-01
## 2400	83	4.835	3.049	4.570e-03	3.543	4.000	2.700e-01
## 2401	2	8.000	4.000	1.000e+00	4.000	8.000	1.000e+00
## 2402	50	4.182	3.315	2.952e-01	3.086	3.951	4.031e-01
## 2403	176	6.737	5.538	2.553e-01	6.492	6.081	9.642e-01
## 2404	148	5.324	3.654	4.334e-03	4.845	4.495	5.939e-01
## 2405	67	5.219	4.024	4.996e-02	4.640	4.530	5.227e-01
## 2406	73	5.759	5.323	7.580e-01	5.623	5.606	7.625e-01
## 2500	273	4.194	3.562	2.646e-02	4.162	3.578	6.545e-02
## 2501	11	2.093	1.516	6.177e-01	1.838	1.782	1.000e+00
## 2502	81	5.161	4.415	3.733e-01	4.167	4.674	4.472e-01
## 2503	47	3.815	3.322	2.980e-01	3.464	3.425	6.641e-01
## 2504	83	4.265	3.919	5.573e-01	4.075	3.970	6.853e-01
## 2505	65	4.432	4.031	8.560e-01	4.713	3.973	4.030e-01
## 2506	70	4.926	3.554	5.702e-02	4.779	3.661	1.111e-01
## 2507	62	3.610	4.242	1.710e-01	4.151	4.009	8.311e-01
## 2508	90	4.383	4.077	8.056e-01	4.493	4.044	4.369e-01
## 2600	90	1.943	1.465	1.528e-02	1.631	1.546	4.764e-01
## 2601	18	2.169	1.546	4.156e-01	2.000	1.644	1.000e+00
## 2602	29	2.828	1.387	6.092e-02	2.289	1.383	6.022e-02
## 2603	26	2.000	1.279	1.669e-01	1.000	1.363	5.264e-01
## 2604	124	2.867	1.866	8.499e-03	2.566	1.895	7.321e-02
## 2605	29	6.055	1.697	2.722e-03	3.000	1.994	4.935e-01
## 2606	35	2.280	2.147	7.867e-01	2.213	2.155	8.710e-01
## 2607	19	2.236	2.006	8.366e-01	2.515	1.916	4.376e-01
## 2608	13	1.000	1.230	7.171e-01	1.732	1.134	2.844e-01
## 2609	NA	NA	NA	NA	NA	NA	NA
## 2610	29	1.817	1.822	1.000e+00	2.000	1.802	7.633e-01
## 2611	74	3.252	2.492	8.633e-02	3.018	2.552	3.315e-01
## 2612	NA	NA	NA	NA	NA	NA	NA
## 2613	104	2.791	1.906	4.294e-02	2.503	1.966	1.722e-01
## 2614	27	4.000	1.939	1.837e-01	3.915	1.830	2.279e-02
## 2700	477	3.761	3.230	1.259e-02	3.500	3.543	6.579e-01
## 2701	324	4.772	3.922	2.967e-02	4.338	4.596	1.671e-01
## 2702	37	3.902	3.264	4.302e-01	3.530	3.415	8.490e-01

## 2703	107	4.339	3.460	5.475e-02	3.876	3.894	9.667e-01
## 2704	15	3.634	5.156	6.622e-01	5.462	4.510	1.944e-01
## 2705	294	5.704	5.013	2.003e-01	5.083	5.261	4.093e-01
## 2706	92	4.344	4.216	8.872e-01	4.152	4.339	8.854e-01
## 2707	61	3.488	3.009	4.133e-01	3.552	2.992	3.005e-01
## 2708	98	4.364	4.597	5.534e-01	4.258	4.575	7.466e-01
## 2709	NA	NA	NA	NA	NA	NA	NA
## 2710	9	3.757	5.000	6.831e-01	3.630	4.213	6.985e-01
## 2711	138	3.783	4.290	2.232e-01	3.508	4.278	1.279e-01
## 2712	165	5.580	4.979	2.786e-01	5.304	5.389	6.989e-01
## 2713	106	5.470	4.282	5.357e-02	5.132	4.758	6.717e-01
## 2714	55	3.091	2.469	2.022e-01	3.029	2.738	4.276e-01
## 2715	93	3.717	4.161	2.942e-01	3.594	4.156	3.387e-01
## 2716	82	7.011	6.470	6.953e-01	6.718	6.860	6.425e-01
## 2717	142	4.126	5.009	4.422e-02	4.146	4.795	1.182e-01
## 2718	137	4.993	3.801	2.105e-03	4.938	4.156	4.589e-02
## 2719	395	3.987	3.173	5.377e-03	3.819	3.561	4.767e-01
## 2720	76	5.224	4.700	7.288e-01	4.950	5.061	6.644e-01
## 2721	40	3.769	3.751	8.534e-01	3.775	3.750	7.172e-01
## 2722	35	5.623	3.450	2.087e-02	4.260	4.443	4.598e-01
## 2723	115	6.819	5.831	3.866e-01	6.292	6.599	3.359e-01
## 2724	166	5.587	4.231	6.959e-02	5.190	4.886	6.662e-01
## 2725	285	5.500	5.071	5.208e-01	5.025	5.434	2.857e-01
## 2726	115	4.940	4.408	5.493e-01	5.112	4.480	3.231e-01
## 2727	61	4.178	4.723	3.772e-01	4.588	4.402	6.373e-01
## 2728	351	5.445	4.458	1.658e-03	4.996	4.855	6.535e-01
## 2729	290	4.198	3.682	1.769e-01	3.915	4.466	4.083e-03
## 2730	487	6.376	4.799	2.109e-12	6.173	5.203	2.815e-05
## 2731	123	4.464	3.651	4.746e-02	4.084	3.997	7.674e-01
## 2732	419	4.112	4.358	5.605e-01	3.908	4.401	6.648e-02
## 2733	81	3.644	4.135	3.807e-01	3.342	4.251	6.927e-02
## 2734	230	3.567	3.062	1.210e-01	3.299	3.337	8.601e-01
## 2735	389	4.405	3.339	4.071e-03	4.324	3.672	1.051e-01
## 2736	230	4.367	3.706	8.272e-02	3.823	4.311	2.496e-01
## 2737	180	4.797	4.377	1.820e-01	4.571	4.484	5.132e-01
## 2738	685	4.178	3.576	1.171e-02	3.789	4.065	4.868e-02
## 2739	907	4.098	3.775	1.093e-01	3.935	4.060	3.603e-01
## 2740	155	5.857	4.930	5.473e-02	5.774	5.112	1.489e-01
## 2741	364	4.569	4.064	9.043e-03	4.330	4.145	1.748e-01
## 2742	135	4.157	3.542	2.838e-01	4.054	3.919	9.449e-01
## 2743	62	4.009	4.395	7.077e-01	3.926	4.307	6.300e-01
## 2744	NA	NA	NA	NA	NA	NA	NA
## 2745	35	6.278	5.011	2.382e-01	5.166	5.770	4.849e-01
## 2746	437	4.314	4.251	7.210e-01	4.405	4.244	4.464e-01
## 2747	29	5.899	5.073	6.877e-01	6.553	5.133	4.471e-01
## 2748	70	4.592	5.248	3.550e-01	5.034	4.994	9.506e-01
## 2800	198	4.447	3.869	2.399e-01	4.282	4.055	7.726e-01
## 2801	25	5.658	5.237	4.546e-01	5.519	5.555	7.795e-01
## 2802	116	4.315	4.176	8.923e-01	3.894	4.564	1.577e-01
## 2803	95	5.217	4.511	2.319e-01	4.449	5.312	1.334e-01

## 2804	105	5.508	4.798	1.816e-01	5.339	5.089	7.081e-01
## 2805	110	3.893	3.823	9.446e-01	3.600	4.090	2.351e-01
## 2806	31	4.665	4.008	3.126e-01	5.377	3.713	4.037e-02
## 2807	18	4.168	5.595	7.879e-01	3.184	5.816	7.179e-02
## 2808	183	6.040	4.451	1.808e-03	5.258	5.094	7.967e-01
## 2809	75	4.646	3.892	2.136e-01	4.098	4.317	4.057e-01
## 2900	307	3.467	3.274	4.737e-01	3.394	3.514	6.288e-01
## 2901	15	3.796	1.000	1.550e-01	3.223	3.651	9.039e-01
## 2902	56	3.477	4.696	2.042e-01	2.553	5.329	6.762e-04
## 2903	3	6.481	4.000	6.667e-01	6.481	4.000	6.667e-01
## 2904	6	5.091	3.464	2.188e-01	6.481	3.722	8.515e-02
## 2905	78	3.700	4.007	4.099e-01	3.705	3.887	8.562e-01
## 2906	27	2.837	5.460	7.165e-03	3.077	4.504	1.617e-01
## 2907	58	3.582	4.656	9.243e-02	3.435	4.709	2.494e-02
## 2908	7	4.863	4.000	7.990e-01	6.649	3.663	4.768e-02
## 2909	49	3.471	4.790	2.595e-01	3.404	4.510	2.800e-01
## 2910	39	2.177	1.503	5.388e-02	2.205	1.592	1.010e-01
## 2911	15	3.917	2.289	2.068e-01	3.910	2.632	3.137e-01
## 2912	33	3.181	2.289	1.410e-01	2.709	4.169	7.163e-02
## 2913	73	3.247	3.000	4.137e-01	3.136	4.409	9.155e-02
## 2914	NA	NA	NA	NA	NA	NA	NA
## 2915	NA	NA	NA	NA	NA	NA	NA
## 2916	274	4.709	4.415	7.031e-01	4.594	4.779	2.906e-01
## 2917	10	4.447	5.000	1.000e+00	3.667	6.117	5.818e-02
## 2918	NA	NA	NA	NA	NA	NA	NA
## 2919	31	3.205	3.557	8.881e-01	3.274	3.000	2.264e-01
## 2920	4	1.260	1.000	1.000e+00	1.000	1.414	6.171e-01
## 2921	94	3.459	3.793	5.029e-01	3.523	3.696	9.185e-01
## 2922	NA	NA	NA	NA	NA	NA	NA
## 2923	13	2.965	3.464	7.518e-01	2.864	3.464	5.640e-01
## 3000	26	3.984	2.573	1.212e-01	3.166	3.662	7.846e-01
## 3001	6	6.309	8.367	8.000e-01	6.415	7.489	1.000e+00
## 3002	98	5.827	4.301	3.039e-02	4.271	4.806	3.977e-01
## 3003	81	4.197	4.601	1.494e-01	3.941	4.750	2.288e-02
## 3004	194	5.243	4.215	2.311e-02	4.525	4.860	5.188e-01
## 3005	88	4.871	3.769	3.106e-02	3.923	4.590	7.735e-02
## 3100	154	4.859	2.594	4.755e-05	3.385	2.940	4.463e-01
## 3101	26	2.852	2.414	9.773e-01	1.732	2.623	4.022e-01
## 3102	25	2.885	2.544	2.775e-01	2.667	2.641	6.383e-01
## 3103	46	2.890	2.362	3.711e-01	2.605	2.427	7.715e-01
## 3104	194	4.039	3.190	1.076e-01	4.050	3.220	1.251e-01
## 3105	61	3.996	4.019	9.117e-01	4.384	3.919	5.476e-01
## 3106	20	2.000	2.127	9.119e-01	8.000	1.965	2.042e-01
## 3107	99	4.242	3.505	3.855e-01	3.597	3.639	9.162e-01
## 3108	28	5.277	3.998	2.286e-01	4.224	4.250	8.705e-01
## 3109	35	2.621	2.277	6.452e-01	3.826	2.197	1.406e-01
## 3110	19	4.380	4.851	7.759e-01	4.738	4.782	9.594e-01
## 3200	281	2.869	2.436	4.436e-02	2.893	2.478	4.108e-02
## 3201	40	2.758	2.172	3.168e-01	2.551	2.459	8.034e-01
## 3202	204	3.001	2.523	5.149e-02	2.718	2.873	4.030e-01

## 3203	309	3.618	3.671	6.874e-01	3.548	3.742	3.811e-01
## 3204	394	3.045	2.567	1.646e-02	3.009	2.740	1.418e-01
## 3205	229	3.129	2.948	5.140e-01	2.904	3.162	2.194e-01
## 3206	97	3.700	3.051	1.514e-01	3.496	3.363	9.678e-01
## 3207	239	2.426	2.132	8.250e-02	2.377	2.236	3.862e-01
## 3300	177	1.737	1.906	3.765e-01	1.797	1.853	7.938e-01
## 3301	246	2.218	1.803	1.168e-02	2.116	1.988	4.800e-01
## 3302	68	2.349	2.103	5.463e-01	2.214	2.272	9.495e-01
## 3303	157	1.941	1.566	2.323e-02	1.733	1.840	5.453e-01
## 3304	1248	2.134	1.916	1.757e-03	2.053	2.064	9.214e-01
## 3305	559	1.834	1.865	7.639e-01	1.790	1.909	2.096e-01
## 3306	360	2.971	2.556	7.570e-02	2.768	2.967	3.756e-01
## 3307	63	3.591	2.855	9.887e-02	3.180	3.114	6.349e-01
## 3308	454	1.775	1.531	6.646e-03	1.626	1.689	6.900e-01
## 3309	88	1.532	2.087	4.129e-02	1.450	2.101	4.596e-03
## 3310	243	1.960	1.532	8.380e-03	1.863	1.686	3.195e-01
## 3311	69	3.153	2.640	2.948e-01	2.560	2.924	3.685e-01
## 3312	960	1.837	1.425	2.760e-11	1.702	1.584	4.604e-02
## 3313	104	3.022	2.507	7.066e-02	2.744	2.640	7.315e-01
## 3314	216	1.472	1.321	6.352e-02	1.392	1.426	9.098e-01
## 3315	216	1.605	1.531	4.560e-01	1.580	1.567	8.346e-01
## 3316	507	1.334	1.223	2.738e-02	1.315	1.244	9.366e-02
## 3317	58	1.514	1.678	4.372e-01	1.642	1.490	5.201e-01
## 3318	109	1.588	1.361	1.987e-01	1.468	1.725	2.641e-01
## 3319	68	2.263	2.371	9.588e-01	2.333	2.245	7.320e-01
## 3320	298	1.220	1.230	5.644e-01	1.212	1.234	2.863e-01
## 3321	115	1.946	1.509	3.065e-02	1.730	1.747	9.618e-01
## 3322	126	2.007	1.767	1.485e-01	1.893	1.881	9.980e-01
## 3400	141	4.047	4.142	7.466e-01	3.781	4.294	3.444e-01
## 3401	10	2.825	2.702	1.000e+00	3.000	2.707	1.000e+00
## 3402	5	5.477	2.289	1.386e-01	5.477	2.289	1.386e-01
## 3403	36	4.922	2.495	2.265e-03	5.293	3.441	2.122e-02
## 3404	8	4.356	2.060	5.451e-02	5.477	2.449	6.030e-02
## 3500	23	3.697	3.887	8.976e-01	2.213	4.274	2.817e-02
## 3501	NA	NA	NA	NA	NA	NA	NA
## 3502	NA	NA	NA	NA	NA	NA	NA
## 3503	NA	NA	NA	NA	NA	NA	NA
## 3504	13	6.000	3.979	2.293e-01	4.000	4.408	8.837e-01
## 3505	3	5.477	3.000	6.667e-01	3.000	5.477	6.667e-01
## 3506	NA	NA	NA	NA	NA	NA	NA
## 3600	9	1.842	6.000	7.151e-02	NA	NA	NA
## 3601	18	2.180	2.768	3.062e-01	2.353	2.376	9.611e-01
## 3602	16	4.268	1.911	2.432e-02	3.022	2.405	6.381e-01
## 3603	6	2.449	6.000	9.018e-02	3.780	2.884	4.936e-01
## 3604	27	3.607	2.652	2.576e-01	3.425	3.024	7.048e-01
## 3605	34	4.627	3.148	8.561e-02	3.327	3.811	3.191e-01
## 3606	NA	NA	NA	NA	NA	NA	NA
## 3607	9	4.481	4.364	7.946e-01	4.000	4.456	6.961e-01
## 3608	NA	NA	NA	NA	NA	NA	NA
## 3609	54	3.692	3.151	6.361e-01	3.782	2.927	2.948e-01

```
## 3610      30  4.499  2.380 4.688e-03 3.958 3.031 3.475e-01
## 3611      80  2.791  2.772 8.956e-01 2.583 3.138 1.395e-01
## 3612     361  4.071  3.851 2.338e-01 3.846 4.050 4.652e-01
## 3613       3  4.000  2.449 6.667e-01 2.000 3.464 6.667e-01
## 3614      53  3.544  3.363 9.372e-01 3.777 3.259 6.147e-01
## 3615      NA      NA      NA      NA      NA      NA      NA
## 3616     106  3.427  3.120 4.790e-01 3.277 3.605 5.182e-01
```

```
print(RegCoef)
```

```
##          FFA1          FLA1          2          3          4          5+
## 1000  0.0179999 -0.0883110  1.039649  1.31658  1.514465  1.5656401
## 1100  0.0155377 -0.0653818  0.331382  0.45912  0.492446  0.5615448
## 1101 -0.0644687 -0.0829115  0.221366  0.23196  0.301566  0.3466751
## 1102  0.0056448 -0.0171104  0.183137  0.25030  0.302961  0.4266906
## 1103  0.0186889 -0.0105605  0.117877  0.16186  0.192764  0.2417738
## 1104  0.0188927  0.0044348  0.068765  0.06717  0.113511  0.1716035
## 1105 -0.0098673 -0.0312695  0.155188  0.19730  0.253954  0.2968933
## 1106  0.0100559 -0.0099816  0.395410  0.45621  0.597609  0.6003958
## 1107 -0.0257242 -0.0136728  0.056854  0.16036  0.223317  0.2059101
## 1108 -0.0842135  0.0042657  0.169733  0.09408  0.222217  0.2998685
## 1109  0.0005040 -0.0311519  0.227331  0.27893  0.364671  0.4203054
## 1110 -0.0419263 -0.0725153  0.213559  0.31687  0.422552  0.5231923
## 1111  0.0262363  0.0450365 -0.003472  0.09045  0.067990  0.0861036
## 1200  0.0725562 -0.2773210  0.557542  0.57638  0.585435  0.5348919
## 1201 -0.0055706  0.0080057  0.190921  0.19349  0.187881  0.2568775
## 1202 -0.0476087  0.0518634  0.195523  0.31718  0.649244  0.1233655
## 1203  0.0362507  0.0212598  0.245452  0.38246  0.441050  0.5183230
## 1204 -0.0486181 -0.0235907  0.291377  0.40433  0.389114  0.2219395
## 1205 -0.0038646 -0.2425327 -0.246124 -0.78120 -0.448699  0.6263013
## 1206 -0.3843474  0.2846397  0.058246  1.02379 -0.013242 -0.2072145
## 1207  0.0849345  0.0589339  0.319094  0.47047  0.624771  0.6146014
## 1208  0.0803321  0.0993336  0.251719  0.14659  0.700066  1.1673844
## 1209 -0.2614514  0.0815222  0.009998  0.97815  1.211013 -0.0354613
## 1210  0.1523097 -0.0592553  0.560586  0.79872  1.151677  1.4855409
## 1211 -0.0104207 -0.0473109  0.145548  0.50812  0.501467  0.4669661
## 1212 -0.1201480  0.1835042  0.316334  0.37713  0.730551  0.4115014
## 1213 -0.0152587  0.1895024  0.301886  0.62240  0.722980  0.2939516
## 1300 -0.0214244 -0.0612388  0.318895  0.23726  0.325910  0.4039854
## 1301      NA      NA      NA      NA      NA      NA
## 1302 -0.0019321 -0.0077216  0.336762  0.28707  0.147631  0.2848750
## 1303 -0.0108223 -0.0042008  0.184670  0.20607  0.228659  0.3115324
## 1304 -0.0089778  0.0076435  0.248897  0.32014  0.350460  0.3462158
## 1305 -0.0455088 -0.0025183  0.481176  0.51338  0.577264  0.6420131
## 1306 -0.0234353 -0.0237730  0.491645  0.62313  0.611591  0.7120869
## 1307 -0.0167628 -0.0447410  0.101853  0.15579  0.187806  0.2667925
## 1308 -0.0188305 -0.0715141  0.300879  0.38153  0.376127  0.5221555
## 1309 -0.0217933 -0.0417705  0.043236  0.05804  0.040028  0.1366760
## 1310  0.0094820 -0.0100361  0.062629  0.11590  0.134230  0.2431520
## 1311 -0.0507122 -0.0527854  0.106035  0.14510  0.201118  0.2602164
```


##	1312	-0.0175265	-0.0460719	0.136752	0.17402	0.202495	0.2939793
##	1313	-0.0030042	0.0282329	0.194802	0.30119	0.319042	0.3970933
##	1314	-0.0152244	-0.0524670	0.042786	0.06805	0.133876	0.1724834
##	1315	-0.0152809	-0.0091346	0.303886	0.30771	0.222188	0.3867895
##	1400	-0.0219266	0.0174996	0.194115	0.41465	0.351463	0.4383086
##	1401	-0.0213877	0.0487036	0.163613	0.09259	0.230423	0.2026727
##	1402	0.0239143	0.1112481	0.137481	0.21935	0.114003	-0.0506623
##	1403	0.0461459	0.0144291	0.114048	0.17208	0.231662	0.2514699
##	1404	0.1283522	0.0258916	0.137189	0.12582	0.396866	0.6307232
##	1405	0.0349115	-0.0222427	0.141166	0.20491	0.310386	-0.4130965
##	1406	0.0328765	0.0451430	0.204515	0.20662	0.100628	0.1600370
##	1407	0.0450607	0.0093651	0.179199	0.22484	0.360297	0.3924572
##	1408	0.0834169	-0.0424114	0.093083	0.21728	0.172654	0.0251006
##	1409	0.0606570	-0.0508654	0.041956	0.09747	0.056945	0.0810731
##	1410	0.1527411	0.0458485	0.115737	0.39381	0.298705	0.3227616
##	1500	0.0303025	0.0783890	0.321609	0.27591	0.275858	0.3228751
##	1501	-0.0927030	2.2033496	0.193853	0.13880	0.756838	0.7173876
##	1502	0.0173330	0.0707696	0.597454	0.65998	0.723277	0.7071489
##	1503	-0.0987246	-0.0662916	-0.087687	0.01376	-0.019743	0.0692827
##	1504	0.0785484	0.2076546	-0.060518	-0.09612	-0.088478	-0.5287081
##	1505	-0.0488197	0.0450756	-0.046895	-0.02416	-0.024832	0.0427169
##	1506	0.0867245	0.1589965	0.486860	0.57345	0.454672	0.7663739
##	1507	-0.0391048	-0.0109839	0.217789	0.24981	0.116547	0.0080762
##	1508	0.2194949	-0.0817789	0.326879	0.60637	0.354754	0.5957430
##	1600	0.0242481	0.0286170	0.133915	0.05781	0.122515	0.2136421
##	1601	NA	NA	NA	NA	NA	NA
##	1602	0.0163179	0.0425086	-0.027456	0.09032	0.114389	0.1439517
##	1603	0.0643795	-0.0205029	0.010308	0.04394	0.157673	0.2042263
##	1604	0.0255158	0.0367365	-0.149574	-0.13328	-0.118138	-0.0610843
##	1605	-0.0183441	0.0322470	0.087207	0.14369	0.161986	0.2063237
##	1606	0.0383154	0.0714127	0.083327	0.11714	0.105568	0.1538632
##	1607	0.0525447	0.0366786	0.019002	0.04419	0.054015	0.0983150
##	1700	0.2254685	-0.0450726	0.118225	0.27480	0.338862	0.2425109
##	1701	-0.4530709	-0.3461126	0.090409	-0.03469	0.879539	0.1661484
##	1702	0.0160275	0.1071816	0.099604	0.19969	0.168600	0.2460712
##	1703	-0.0722442	-0.0521943	0.188455	0.38495	0.467831	0.4528531
##	1704	-0.0368759	-0.2244325	0.590550	0.62535	0.441345	0.1967219
##	1705	-0.1195384	0.0303580	0.007725	0.31150	0.344209	0.4017983
##	1706	-0.0436504	-0.0641959	0.257294	0.33565	0.357770	0.3670450
##	1707	-0.1263439	-0.0922164	0.049898	0.18837	0.028758	0.2949430
##	1708	-0.0036012	0.0534932	0.103936	0.37239	0.548930	0.4744508
##	1709	0.0887607	0.0632428	0.245560	0.42411	0.385078	0.5185905
##	1710	0.0093111	-0.0166144	0.241835	0.35846	0.414033	0.3137852
##	1711	0.0036670	-0.0324013	-0.037988	0.01645	0.293025	0.0950931
##	1712	-0.0252391	0.0141863	0.187464	0.36532	0.464925	0.3878119
##	1800	0.0615402	0.0799244	0.163271	0.17450	0.261390	0.1892239
##	1801	NA	NA	NA	NA	NA	NA
##	1802	0.0317667	-0.0811879	0.130810	0.25986	0.209778	0.4602382
##	1803	0.0631574	0.0196681	0.057265	0.04796	0.206473	0.1837539
##	1804	0.0883970	0.0074391	0.254462	0.27207	0.488174	-0.1426104

##	1900	-0.1195141	0.0697847	0.263865	0.36263	0.198118	0.3569861
##	1901	-0.0535261	0.0925981	0.076861	0.16577	0.182910	0.1727164
##	1902	-0.0416213	0.0503341	0.152310	0.23447	0.221660	0.3547055
##	1903	-0.1379110	0.1032811	0.110745	0.30421	0.118450	0.4141801
##	1904	0.0086380	-0.0025728	0.078912	0.13039	0.095709	0.1189186
##	1905	-0.0760493	0.0703480	0.106904	0.17697	0.034355	0.2059114
##	1906	-0.0639417	-0.0123443	0.107906	0.16253	0.142216	0.2290075
##	1907	-0.0550266	-0.0211074	0.194807	0.31151	0.335621	0.3721992
##	1908	-0.0965822	0.0483497	0.098890	0.23924	0.278708	0.3036484
##	1909	0.0388540	-0.1337448	0.363549	0.76257	0.566073	0.6997440
##	1910	-0.0236085	0.0176922	0.121327	0.12430	0.145107	0.2383603
##	1911	0.0777208	-0.1053470	0.077244	0.18635	0.269245	0.1830417
##	1912	-0.0993572	-0.0099559	0.210565	0.28016	0.277734	0.2070974
##	1913	-0.0811845	-0.0393081	0.230045	0.26148	0.255120	0.1713958
##	2000	0.0045492	0.0888769	0.122916	0.17590	0.294424	0.4453749
##	2001	-0.0359050	0.2190310	0.189346	0.32361	0.350206	-0.3718090
##	2002	-0.0069139	0.0652442	0.135961	0.20982	0.255518	0.4790712
##	2003	-0.0105033	0.1325424	0.108732	0.16451	0.043906	-0.0957630
##	2100	NA	NA	NA	NA	NA	NA
##	2101	NA	NA	NA	NA	NA	NA
##	2102	0.0664787	-0.1006105	0.510592	0.66000	0.599484	0.3525105
##	2103	0.0207967	-0.0691659	0.465214	0.56390	0.480869	0.2971211
##	2104	0.0548687	-0.4085977	0.212362	0.68583	0.435140	0.6619260
##	2105	0.0511987	-0.1372926	0.180815	0.25801	0.373321	0.2600294
##	2200	NA	NA	NA	NA	NA	NA
##	2201	0.0616309	-0.1509204	0.165240	0.61935	0.436511	0.8619394
##	2202	-0.1074571	-0.0064220	0.321403	0.54347	0.381053	0.2953619
##	2203	0.0234077	-0.1154532	0.291389	0.55456	0.432705	0.0136904
##	2204	-0.0202917	-0.0076068	0.153168	0.18098	0.301286	0.2013060
##	2205	0.0074019	-0.0429404	0.196656	0.23976	0.244212	0.3900318
##	2206	0.3699125	-0.5699539	0.252182	0.25378	0.148028	0.3954835
##	2207	0.0033950	-0.1733881	0.201079	0.32186	0.293428	0.1022066
##	2208	-0.0170856	0.0114679	0.114788	0.09906	0.107301	0.0758728
##	2209	0.0854074	-0.0449480	0.126836	0.21562	0.406159	0.3222815
##	2210	0.0922633	0.0986191	0.202478	0.33934	0.352852	0.4520970
##	2211	0.0473686	-0.0210763	0.203556	0.27626	0.291730	0.3782915
##	2212	-0.0699253	-0.0757471	0.036606	0.12931	0.002227	0.2102185
##	2213	0.1055608	0.0972812	0.080399	0.29073	0.258181	0.3171251
##	2214	-0.0810831	-0.0573049	0.504620	0.41807	0.450842	0.7599531
##	2215	0.1059472	0.0593052	0.253125	0.25905	0.190624	0.5558271
##	2216	-0.3464079	0.2788829	0.089643	0.25484	0.234919	-0.2670239
##	2300	0.0165541	-0.0622144	0.185866	0.24894	0.336120	0.4141092
##	2301	0.0136572	0.0089852	0.160787	0.12667	0.250114	-0.0004539
##	2302	0.0962951	0.1193172	0.099228	0.20187	0.305711	0.3289368
##	2303	-0.0154965	-0.0055844	0.155437	0.16941	0.213310	0.2786317
##	2304	-0.0158610	0.0792563	0.222484	0.17598	0.158963	0.2778067
##	2305	0.0905263	0.0217987	0.317426	0.34914	0.320256	0.3961674
##	2306	0.0187093	0.1097203	0.155046	0.17029	0.277006	0.2457327
##	2307	-0.0313269	-0.0373526	0.080527	0.02163	0.139200	0.1840130
##	2308	-0.0395295	-0.0148436	0.121027	0.16757	0.204028	0.2641237

##	2309	-0.0375123	-0.0369020	0.117432	0.19951	0.295378	0.2695299
##	2310	0.0251391	0.0438067	0.219241	0.29048	0.265137	0.3769670
##	2311	0.0782825	0.0619699	0.275039	0.30819	0.411302	0.3894675
##	2312	0.0077571	0.0064056	0.297604	0.38786	0.300337	0.3864784
##	2400	-0.0797531	0.0345730	0.137294	0.16351	0.175494	0.1049222
##	2401	NA	NA	NA	NA	NA	NA
##	2402	0.0103106	0.0321968	0.732818	0.78543	0.796233	0.9289492
##	2403	-0.0368517	-0.0365954	0.102500	0.14595	0.188463	0.2887614
##	2404	-0.0254560	-0.0282201	0.171611	0.26940	0.295063	0.3383342
##	2405	-0.0446402	0.0107500	0.171530	0.21697	0.295288	0.3985929
##	2406	-0.0360583	-0.0053065	0.066279	0.03859	0.111217	0.1354969
##	2500	0.1171624	0.0227672	0.251551	0.09720	0.222671	0.3987516
##	2501	0.0113041	0.1761031	0.697871	0.69609	0.638628	0.6791702
##	2502	-0.0087761	0.0273506	0.159570	0.19690	0.328145	0.3229532
##	2503	-0.0519730	-0.0591580	0.451735	0.51404	0.581834	0.5979477
##	2504	-0.0693401	-0.0077875	-0.017445	0.04069	0.113795	0.1368030
##	2505	0.0281678	0.0171575	0.062942	0.07111	0.067442	0.1044710
##	2506	0.0197395	0.0832060	0.271280	0.32289	0.356012	0.3050615
##	2507	0.0164096	-0.0250747	0.561440	0.63117	0.658932	0.6784784
##	2508	-0.0500647	-0.0389248	0.017903	0.07606	0.102793	0.1223283
##	2600	-0.0330640	-0.0487825	0.342089	0.27216	0.246493	0.4187430
##	2601	0.1053299	0.2054630	0.105815	0.17903	0.346337	-0.0358736
##	2602	-0.1268549	-0.0554974	0.264902	-0.02224	0.114855	-0.0049504
##	2603	-0.2449222	-0.1375308	0.269988	-0.09592	0.191764	-0.0474762
##	2604	-0.0174598	-0.0714758	0.206277	0.19539	0.299484	0.2176678
##	2605	-0.2325867	0.1056467	0.116200	0.09641	0.190675	0.3403726
##	2606	0.2732068	-0.1573252	0.152099	-0.07053	0.229640	0.1472264
##	2607	-0.0615672	-0.1554203	0.080159	0.15932	0.183290	0.0853667
##	2608	-0.3693392	0.1811284	0.252727	0.08004	0.822621	-0.3652906
##	2609	0.3706670	-0.6337501	0.458022	-0.69960	-0.283776	0.7699744
##	2610	-0.1586866	-0.0898508	0.009400	0.32170	0.219786	-1.3462681
##	2611	0.0056270	0.0264759	0.239518	0.30197	0.295180	0.3848807
##	2612	-0.2423649	0.1966184	0.279105	0.29069	0.351868	-0.6737851
##	2613	-0.0767694	-0.0278490	0.206770	0.38301	0.458477	0.3091468
##	2614	0.1397489	-0.0836586	0.050899	0.26911	0.289055	0.6379576
##	2700	0.0268011	0.0226042	0.580778	0.88743	1.039425	1.3009303
##	2701	0.0560759	-0.0027501	0.218677	0.30938	0.336747	0.4390720
##	2702	-0.1082300	0.0056854	0.127933	0.20665	0.319269	0.4538033
##	2703	0.1112034	0.0578401	0.480437	0.48364	0.589471	0.7504537
##	2704	-0.0057367	-0.0476483	0.559660	0.79610	0.717310	1.0475956
##	2705	0.0793761	0.1253186	0.188697	0.29398	0.375333	0.5990297
##	2706	0.0432007	0.1037883	0.369394	0.53781	0.614323	0.9294973
##	2707	0.1322977	-0.1247256	0.620652	0.64776	0.723039	0.7911384
##	2708	-0.0079372	-0.0074444	0.138333	0.24203	0.307628	0.3688785
##	2709	NA	NA	NA	NA	NA	NA
##	2710	-0.0761753	0.0075347	0.038152	0.04416	-0.020365	0.0712709
##	2711	0.0838664	0.0762580	0.304836	0.44915	0.526159	0.6551094
##	2712	0.0109461	-0.0337036	0.113346	0.27361	0.286487	0.4465356
##	2713	0.0590130	-0.0123187	0.215435	0.33493	0.325524	0.3432850
##	2714	0.0419926	0.0113998	0.238697	0.19915	0.424139	0.5153909

##	2715	-0.0058460	0.0440524	0.248715	0.29311	0.305245	0.4236017
##	2716	-0.0828007	-0.0411562	0.028830	0.10946	0.171478	0.1945953
##	2717	-0.0459859	0.0199320	0.233451	0.25472	0.307516	0.3997026
##	2718	0.1611690	0.0173710	0.168003	0.22043	0.262096	0.3283127
##	2719	0.0135600	-0.0108098	0.280239	0.31303	0.395223	0.4662283
##	2720	-0.0487138	-0.0620644	0.286648	0.42267	0.376706	0.6061029
##	2721	-0.0303055	-0.0349732	0.289509	0.25277	0.334746	0.4666713
##	2722	-0.0143840	-0.0086813	-0.011404	0.04776	0.164247	0.3685672
##	2723	-0.0630382	-0.0336943	0.489955	0.47241	0.570054	0.6635279
##	2724	0.0026534	-0.0696748	0.235538	0.45402	0.391312	0.6272440
##	2725	-0.0102562	0.0201362	0.123309	0.14363	0.237223	0.2915148
##	2726	0.0162510	-0.0068254	0.736106	0.71358	0.823498	0.8827434
##	2727	-0.0400222	-0.0381320	0.186098	0.25343	0.259331	0.4421897
##	2728	0.0647003	0.0611776	0.212190	0.26979	0.311031	0.4145514
##	2729	0.1204821	0.0955784	0.216024	0.21326	0.231393	0.3171992
##	2730	0.0009667	0.0112879	0.386303	0.51729	0.505084	0.6113096
##	2731	-0.0293068	-0.0423373	0.311219	0.38380	0.477102	0.5905206
##	2732	-0.0015048	0.0268538	0.089266	0.15257	0.218653	0.2630145
##	2733	0.0416528	0.0614568	0.195767	0.23310	0.389530	0.4012368
##	2734	0.0267128	0.0025965	0.270105	0.31743	0.319557	0.4206595
##	2735	0.0721695	0.0871538	0.243530	0.26979	0.324454	0.4193289
##	2736	0.0190097	-0.0237840	0.401588	0.55808	0.646677	0.7037902
##	2737	0.0470377	-0.0064920	0.210840	0.22083	0.286652	0.4012458
##	2738	0.0615772	0.0262715	0.395770	0.47101	0.519382	0.5706559
##	2739	0.0357094	0.0072659	0.183261	0.27297	0.317454	0.3630342
##	2740	0.0922270	0.0726477	0.344858	0.34612	0.450899	0.6829559
##	2741	-0.0880435	-0.0604212	0.212135	0.41797	0.424096	0.6465673
##	2742	0.0446665	0.0297896	0.328308	0.48982	0.587417	0.6333574
##	2743	0.0462595	-0.0400077	0.224202	0.10765	0.088489	0.1645520
##	2744	NA	NA	NA	NA	NA	NA
##	2745	0.0218305	0.0645737	-0.008228	0.23258	0.350126	0.4038053
##	2746	0.0622428	0.0826553	0.307738	0.35756	0.418192	0.5901631
##	2747	-0.0689409	-0.1236677	0.527168	0.61692	0.634295	0.8134575
##	2748	-0.0107519	0.0109443	0.212430	0.43056	0.457994	0.5617381
##	2800	0.0075172	-0.0203740	0.248634	0.26526	0.323515	0.4064542
##	2801	0.0502077	-0.0195785	-0.087618	0.03492	-0.036383	0.2115205
##	2802	-0.0160967	-0.0301771	0.144836	0.14750	0.167134	0.2647723
##	2803	0.0169916	0.1183366	0.147645	0.26012	0.331286	0.3749792
##	2804	-0.0090416	-0.0384181	-0.067450	0.03076	0.034028	0.0881480
##	2805	-0.0385375	-0.0435798	0.279296	0.30621	0.319572	0.3845176
##	2806	0.0607442	0.0752764	0.260773	0.32012	0.278116	0.3705373
##	2807	-0.0068566	-0.0715445	0.253750	0.22318	0.165902	0.3709105
##	2808	0.0805440	0.0357259	0.194791	0.23338	0.290443	0.3815826
##	2809	-0.0290686	0.0038512	0.160067	0.15566	0.275447	0.3510596
##	2900	0.0084892	0.1189169	0.258077	0.36422	0.430596	0.4650656
##	2901	-0.5319957	-0.2059884	0.747021	0.50413	0.170026	0.8562930
##	2902	-0.0470342	-0.0669360	0.172259	0.38400	0.483572	0.6428518
##	2903	NA	NA	NA	NA	NA	NA
##	2904	NA	NA	NA	NA	NA	NA
##	2905	0.0088234	0.0744676	0.027147	0.14787	0.140976	0.1874939

##	2906	0.1785088	-0.0770737	1.556869	1.33821	1.634778	1.3753217
##	2907	0.1810424	-0.0152986	0.745108	0.78063	0.999968	1.0312519
##	2908	NA	NA	NA	NA	NA	NA
##	2909	0.0837678	0.0250854	0.373072	0.41521	0.415580	0.5268012
##	2910	0.0643358	-0.0379323	0.197462	0.55496	0.470647	0.2056531
##	2911	-0.0083085	-0.0049456	0.101993	0.55547	0.642112	0.9236838
##	2912	0.2395168	-0.0281070	0.242692	0.36011	0.470155	0.2494696
##	2913	0.1796065	0.2405771	0.374381	0.48080	0.677775	0.7583241
##	2914	-0.0962825	0.2265089	0.334366	0.33576	-0.001526	0.5943032
##	2915	NA	NA	NA	NA	NA	NA
##	2916	-0.0239703	-0.0652716	0.359264	0.51781	0.477645	0.5850428
##	2917	0.4608527	0.2002338	0.025200	0.65233	0.166861	0.5179708
##	2918	NA	NA	NA	NA	NA	NA
##	2919	-0.0221166	-0.0444024	0.283858	0.25081	0.212600	0.5006258
##	2920	NA	NA	NA	NA	NA	NA
##	2921	-0.0134171	0.1026968	0.241029	0.19248	0.070212	0.0941895
##	2922	0.2556957	0.0405409	0.069564	0.01743	0.140512	-0.1113523
##	2923	0.2477987	-0.2039803	-0.019654	0.37907	0.167326	0.1573401
##	3000	0.0017567	-0.0772526	0.334689	0.52298	0.556023	0.6398592
##	3001	NA	NA	NA	NA	NA	NA
##	3002	0.0139048	0.0225500	0.084090	0.14823	0.149044	0.2520157
##	3003	0.0002501	-0.0768685	0.266047	0.53549	0.640248	0.5843319
##	3004	-0.0064208	-0.0062898	0.456693	0.58658	0.610837	0.6276625
##	3005	-0.0337125	0.0428996	0.040816	0.15461	0.134454	0.1679157
##	3100	-0.0342783	0.0291240	0.222080	0.33827	0.352251	0.4831859
##	3101	0.0432368	0.0661460	0.122888	0.16447	0.270007	0.3261892
##	3102	-0.1134916	0.1946887	0.298436	0.24819	0.487329	0.5585673
##	3103	-0.0815601	0.0268618	0.308721	0.33579	0.314216	0.1961622
##	3104	0.0478488	-0.0227628	0.093290	0.04088	0.056673	0.1427800
##	3105	-0.0412033	0.0036222	0.046657	0.09317	0.090541	0.0975959
##	3106	-0.0016357	0.0617560	0.086561	0.19762	-0.043531	0.2344874
##	3107	-0.0255695	0.0081399	0.036361	0.09508	0.176890	0.3191301
##	3108	-0.1010040	0.0236790	-0.290016	-0.10694	-0.047797	0.1296031
##	3109	-0.2070123	-0.0410749	0.043329	0.28072	0.100130	0.0722233
##	3110	0.0316174	-0.0259447	0.176355	0.18160	0.275010	0.3094804
##	3200	-0.0072648	0.0688551	0.194441	0.20644	0.200542	0.2863329
##	3201	0.1269537	-0.0211519	0.207877	0.13000	0.309238	0.2191550
##	3202	0.0435449	0.0191993	0.137634	0.18990	0.225072	0.1742167
##	3203	0.0133147	0.0006385	0.280614	0.26977	0.368851	0.4460949
##	3204	-0.0146284	-0.0489085	0.145260	0.18129	0.258996	0.3221746
##	3205	0.0409322	0.0175109	0.132254	0.12259	0.213867	0.2301087
##	3206	0.0196644	0.0450703	0.392138	0.31865	0.423337	0.3996325
##	3207	0.0629886	-0.0512322	0.225990	0.24425	0.247390	0.3231135
##	3300	-0.0634388	-0.0130509	0.186201	0.21829	0.346609	0.2298233
##	3301	-0.0369989	0.0925715	0.113528	0.17135	0.087793	0.1559460
##	3302	-0.1005033	-0.0583143	0.229924	0.41665	0.402055	0.2524957
##	3303	0.0825319	-0.0258469	0.209861	0.22930	0.066359	-0.0301509
##	3304	0.0046205	0.0128336	0.134063	0.20961	0.229541	0.2641255
##	3305	0.0534604	0.0159029	0.144774	0.16276	0.139879	0.2532620
##	3306	0.0758604	-0.0036167	0.148808	0.17993	0.246817	0.2481541

##	3307	0.1021677	0.0801147	0.216479	0.38173	0.429420	0.4928646
##	3308	0.0234811	-0.0307460	0.229597	0.41336	0.456014	0.5691410
##	3309	-0.0236663	-0.0015311	0.378146	0.36324	0.678183	0.6288729
##	3310	0.0146844	0.0121156	0.240079	0.40260	0.418844	0.5639948
##	3311	-0.0267917	0.1883840	0.322006	0.31258	0.299936	0.1633252
##	3312	-0.0092449	0.0340200	0.169888	0.26751	0.346382	0.2159490
##	3313	0.0540559	0.0481857	0.141883	0.17074	0.127111	0.0994633
##	3314	-0.0081504	0.0040363	0.193608	0.28712	0.425686	0.0942303
##	3315	0.1055901	-0.0445079	0.253218	0.56764	0.336746	0.6349790
##	3316	-0.0365776	0.0773999	0.169225	0.28411	0.249602	0.5478169
##	3317	0.1057515	0.0646543	0.121400	0.15478	0.569278	0.1351523
##	3318	0.0332711	-0.0298518	0.333347	0.33327	0.450682	0.4567370
##	3319	-0.0317353	0.0845080	0.184973	0.27925	0.262573	0.1665874
##	3320	0.0731154	0.0099727	0.151063	0.06573	0.404735	-0.4214186
##	3321	-0.0405643	0.1452056	0.031828	0.10669	-0.081639	-0.1593741
##	3322	0.0699764	-0.0904579	0.193169	0.20626	-0.059674	-0.0840211
##	3400	0.0441426	0.0753106	0.245277	0.26374	0.289342	0.2535236
##	3401	NA	NA	NA	NA	NA	NA
##	3402	NA	NA	NA	NA	NA	NA
##	3403	0.1756137	0.1785117	-0.115743	-0.20788	-0.167423	-0.0334738
##	3404	-0.1248831	0.0994863	0.670538	0.51549	0.410383	0.8415678
##	3500	-0.0462132	0.0192544	0.159979	0.36115	0.451251	0.3791019
##	3501	NA	NA	NA	NA	NA	NA
##	3502	NA	NA	NA	NA	NA	NA
##	3503	NA	NA	NA	NA	NA	NA
##	3504	-0.0327425	0.1540575	0.118793	0.10279	0.039880	-0.0363247
##	3505	NA	NA	NA	NA	NA	NA
##	3506	NA	NA	NA	NA	NA	NA
##	3600	0.1655416	-0.0678947	0.188638	0.20532	0.346108	0.3368581
##	3601	0.0702397	-0.0239253	0.323403	0.34290	0.268237	0.1494277
##	3602	-0.0822570	-0.2477057	0.474523	0.58854	1.326654	0.6045563
##	3603	0.2171517	-0.1193002	0.484539	0.37988	0.724917	0.8198495
##	3604	0.0332712	0.0937509	0.276083	0.30932	0.176090	0.1098500
##	3605	-0.1631814	0.2219990	0.339339	0.11692	0.800817	0.2848929
##	3606	NA	NA	NA	NA	NA	NA
##	3607	0.0512253	-0.1136658	0.658453	0.62644	0.544294	0.8219190
##	3608	NA	NA	NA	NA	NA	NA
##	3609	0.1911818	-0.0041487	0.196199	0.16787	0.173325	0.1862864
##	3610	0.0670025	-0.0033340	0.163855	0.39148	0.522000	0.5911469
##	3611	0.0871256	0.0789435	0.494451	0.54740	0.531624	0.7232474
##	3612	-0.0447544	-0.0110903	0.214636	0.35005	0.428254	0.5154714
##	3613	0.0666114	0.0378137	0.274922	0.25730	0.490038	0.4251670
##	3614	-0.0849794	-0.0064105	-0.152865	0.07002	-0.034824	0.3085739
##	3615	NA	NA	NA	NA	NA	NA
##	3616	0.0006407	-0.0527231	0.235712	0.30099	0.419129	0.2999348
##		FFA2	FLA2	FFA3	FLA4		
##	1000	0.3652312	-0.1498681	0.3451204	-0.0574398		
##	1100	0.0667849	-0.0661827	0.0555401	-0.0508667		
##	1101	-0.0178901	-0.0896687	-0.0315043	-0.0934547		
##	1102	0.0416018	-0.0290136	0.0381936	-0.0225061		

## 1103	0.0438633	-0.0227449	0.0387404	-0.0100115
## 1104	0.0225935	0.0036347	0.0231061	0.0076899
## 1105	0.0059908	-0.0446820	-0.0028797	-0.0431326
## 1106	0.0566898	-0.0164552	0.0540060	-0.0042507
## 1107	-0.0121692	-0.0244195	-0.0173979	-0.0279019
## 1108	-0.0595541	-0.0135038	-0.0607882	-0.0193909
## 1109	0.0397250	-0.0408172	0.0334005	-0.0327279
## 1110	-0.0121008	-0.0961210	-0.0261191	-0.0985487
## 1111	0.0346113	0.0481826	0.0387778	0.0527597
## 1200	0.2016217	-0.3921126	-0.1004639	-0.2410441
## 1201	0.0216623	-0.0002181	0.0215642	0.0099213
## 1202	0.0195561	-0.0182445	0.0032067	-0.0005716
## 1203	0.1671165	-0.0434100	0.1356392	0.0756663
## 1204	0.0009406	-0.0813641	-0.0464052	NA
## 1205	0.2510738	-0.4824138	-0.2155904	-0.2445606
## 1206	-0.3384038	0.2544107	-0.1405294	0.0025858
## 1207	0.1557682	0.0938241	0.2229567	0.2079188
## 1208	0.1797896	0.0041572	0.1837362	0.1721959
## 1209	-0.4407964	0.2466264	-0.2537859	-0.0787166
## 1210	0.3972236	-0.1965446	0.2471420	0.1101153
## 1211	0.0219335	-0.0581057	-0.0290319	-0.0388844
## 1212	-0.0434658	0.1235820	0.0630205	0.0855956
## 1213	0.0179594	0.1573283	0.1517868	0.1723316
## 1300	-0.0014598	-0.0547497	-0.0094396	-0.0550343
## 1301	NA	NA	NA	NA
## 1302	0.0072632	0.0113802	0.0101217	0.0130829
## 1303	0.0002924	-0.0022209	0.0002316	-0.0022063
## 1304	0.0142415	0.0147606	0.0155045	0.0168165
## 1305	-0.0233151	-0.0233291	-0.0255960	-0.0267323
## 1306	-0.0086048	-0.0343052	-0.0117868	-0.0352936
## 1307	-0.0061838	-0.0518349	-0.0089332	-0.0523114
## 1308	-0.0036940	-0.0651985	-0.0073432	-0.0654710
## 1309	-0.0197659	-0.0499701	-0.0266564	-0.0531803
## 1310	0.0164475	-0.0187291	0.0155481	-0.0176997
## 1311	-0.0338494	-0.0568047	-0.0384097	-0.0605933
## 1312	-0.0084593	-0.0494467	-0.0106747	-0.0500525
## 1313	0.0014384	0.0184748	0.0020461	0.0185897
## 1314	-0.0062370	-0.0505979	-0.0108679	-0.0514194
## 1315	0.0111020	0.0090164	0.0116966	0.0101519
## 1400	0.0104414	0.0002595	0.0105951	0.0065333
## 1401	-0.0012612	0.0445614	0.0199117	0.0439845
## 1402	0.0386278	0.1032114	0.0869007	0.1225508
## 1403	0.0493003	0.0137761	0.0568004	0.0412320
## 1404	0.1357847	-0.0074903	0.1320664	0.0656033
## 1405	0.0523955	-0.0366002	0.0341094	-0.0083888
## 1406	0.0510918	0.0526211	0.0733107	0.0747989
## 1407	0.0561464	0.0089958	0.0607258	0.0377194
## 1408	0.0966558	-0.0519660	0.0708898	-0.0005923
## 1409	0.0620743	-0.0535994	0.0378947	-0.0249469
## 1410	0.1554408	0.0506248	0.1891457	0.1554878

##	1500	0.0708393	0.0942660	0.0816876	0.1059177
##	1501	-0.1092890	0.1997064	-0.1104704	1.1493976
##	1502	0.0462123	0.0361159	0.0524456	0.0493963
##	1503	-0.0877025	-0.0638222	-0.0897441	-0.0700296
##	1504	0.0078587	0.2007631	0.0209622	0.2058418
##	1505	NA	NA	-0.0492748	0.0479116
##	1506	0.1909834	0.1881085	0.2170170	0.2162296
##	1507	-0.0781701	0.0014639	-0.0808008	-0.0315960
##	1508	0.1921912	-0.1325909	0.1754869	-0.0943055
##	1600	0.0363445	0.0230722	0.0375661	0.0260181
##	1601	0.0059562	0.2633188	0.0442964	0.2623079
##	1602	0.0291399	0.0369432	0.0287814	0.0364614
##	1603	0.0483610	0.0132381	0.0499488	0.0226194
##	1604	0.0239106	0.0342131	0.0249816	0.0354356
##	1605	-0.0153707	0.0366290	-0.0150637	0.0363689
##	1606	0.0475559	0.0755313	0.0512572	0.0795344
##	1607	0.0578240	0.0329920	0.0570781	0.0311488
##	1700	0.2496963	-0.0581277	0.2231293	0.0726412
##	1701	-0.5713733	-0.2382724	-0.7189864	-0.5213655
##	1702	0.0117520	0.1052825	0.0579416	0.1104756
##	1703	-0.0657842	0.0005718	-0.0655836	-0.0200468
##	1704	-0.0223183	-0.0975278	-0.0669711	-0.1067394
##	1705	-0.1390496	0.0246773	-0.1259606	-0.0481401
##	1706	-0.0492108	-0.1032607	-0.0936051	-0.1266069
##	1707	-0.0761944	-0.0629054	-0.0933205	-0.0969439
##	1708	-0.0069758	0.1013687	0.0263447	0.0988388
##	1709	0.0897804	0.0489820	0.1096636	0.0867203
##	1710	0.0250253	-0.0271570	0.0112860	-0.0137359
##	1711	0.0250708	-0.0573941	0.0129558	-0.0515622
##	1712	-0.0108020	0.0243558	-0.0030704	0.0205545
##	1800	0.0756314	0.0587837	0.1040973	0.0994465
##	1801	NA	NA	NA	NA
##	1802	0.0373743	-0.0986183	-0.0143813	-0.0787097
##	1803	0.0638577	0.0224269	0.0751851	0.0563682
##	1804	0.1045554	0.0744852	0.1310247	0.1079640
##	1900	-0.0797656	0.0092353	-0.0761941	-0.0305282
##	1901	-0.0335949	0.0924117	-0.0081102	0.0786174
##	1902	-0.0068669	0.0378935	0.0011262	0.0356380
##	1903	-0.2153759	0.1956233	-0.1190710	0.0774452
##	1904	0.0117629	-0.0185380	0.0050069	-0.0136180
##	1905	-0.0800940	0.1069550	-0.0697533	0.0991639
##	1906	-0.0506924	-0.0005077	-0.0507680	-0.0090369
##	1907	-0.0335756	-0.0099054	-0.0358272	-0.0184096
##	1908	-0.0826182	0.0550595	-0.0727518	0.0354101
##	1909	0.2058929	-0.0944779	0.1924027	-0.0388211
##	1910	-0.0141378	0.0159061	-0.0105438	0.0121322
##	1911	0.0890374	-0.1161053	0.0316296	-0.0671317
##	1912	-0.0697877	-0.0149806	-0.0733737	-0.0387686
##	1913	-0.0525441	-0.0072993	-0.0529472	-0.0146982
##	2000	0.0228544	0.1053647	0.0916804	0.1210583

##	2001	-0.0297862	0.2201401	0.0642049	0.2056326
##	2002	0.0067709	0.0800323	0.0495981	0.0837607
##	2003	0.0034455	0.1332334	0.0616691	0.1348811
##	2100	NA	NA	NA	NA
##	2101	0.2550389	0.0637585	0.3093026	0.2054252
##	2102	0.1297765	-0.2242440	0.0687554	-0.1665089
##	2103	0.0656448	-0.0717712	0.0549172	-0.0559901
##	2104	0.1523203	-0.6295503	-0.1010365	-0.5355212
##	2105	0.0942887	-0.1965911	0.0195701	-0.1486483
##	2200	0.0967684	-0.0645346	0.0756936	-0.0252501
##	2201	0.1118310	-0.2090488	0.0311492	-0.1460118
##	2202	-0.0839725	-0.0649945	-0.1108548	-0.1065055
##	2203	0.0656825	-0.1482623	0.0282536	-0.1208143
##	2204	-0.0085807	-0.0190007	-0.0134413	-0.0222472
##	2205	0.0460355	-0.0673420	0.0252250	-0.0485097
##	2206	0.4074249	-0.5058930	0.4067046	-0.5123704
##	2207	0.0276013	-0.1896544	-0.0218292	-0.1799991
##	2208	-0.0040645	0.0147349	-0.0012075	0.0137483
##	2209	0.1227844	-0.0339622	0.1149895	0.0030429
##	2210	0.1478256	0.1176592	0.1591655	0.1379767
##	2211	0.1075421	-0.0046724	0.1072006	0.0067771
##	2212	-0.0934878	-0.0751113	-0.1233537	-0.1124521
##	2213	0.1431869	0.0927354	0.1822245	0.1598180
##	2214	0.0122866	-0.1521891	-0.0630735	-0.1451060
##	2215	0.1388157	0.0202644	0.1468797	0.0880170
##	2216	-0.3146045	0.2382137	-0.1798654	0.0298554
##	2300	0.0263973	-0.0869745	-0.0056068	-0.0757246
##	2301	0.0409989	-0.0353477	0.0190268	-0.0083849
##	2302	0.1085572	0.1238282	0.1201993	0.1374391
##	2303	-0.0076773	-0.0131941	-0.0101527	-0.0150152
##	2304	-0.0085811	0.0735322	0.0012498	0.0719971
##	2305	0.1089713	-0.0007584	0.1087426	0.0331011
##	2306	0.0234033	0.1190918	0.0553466	0.1257015
##	2307	-0.0183136	-0.0408099	-0.0219340	-0.0431092
##	2308	-0.0297100	-0.0367995	-0.0459583	-0.0510512
##	2309	-0.0242260	-0.0542707	-0.0363646	-0.0607330
##	2310	0.0376743	0.0330250	0.0419892	0.0389039
##	2311	0.0875691	0.0378489	0.0928171	0.0544433
##	2312	0.0395240	-0.0021677	0.0391469	0.0066666
##	2400	-0.0719748	0.0220716	-0.0693921	0.0109865
##	2401	NA	NA	NA	NA
##	2402	0.0406081	0.0194947	0.0420416	0.0242372
##	2403	-0.0324698	-0.0460395	-0.0351575	-0.0487378
##	2404	-0.0167425	-0.0388653	-0.0178929	-0.0396620
##	2405	-0.0464493	-0.0165788	-0.0489175	-0.0266469
##	2406	-0.0330917	-0.0081512	-0.0337719	-0.0125786
##	2500	0.1460595	0.0342544	0.1489383	0.0499395
##	2501	0.0457293	0.1138872	NA	0.1227300
##	2502	0.0076943	0.0313430	0.0112306	0.0326354
##	2503	-0.0014099	-0.0604383	-0.0034922	-0.0594857

##	2504	-0.0585790	0.0055030	-0.0584363	0.0028888
##	2505	0.0320885	0.0166007	0.0335414	0.0209411
##	2506	0.0233215	0.0654395	0.0321277	0.0699708
##	2507	0.0590923	-0.0375707	0.0532922	-0.0216339
##	2508	-0.0431955	-0.0325124	-0.0443335	-0.0352677
##	2600	0.0022410	-0.0311785	-0.0157922	-0.0298308
##	2601	0.1638529	0.1484974	0.2411351	0.2541995
##	2602	-0.0405430	-0.0789391	-0.0807417	-0.1025495
##	2603	-0.1993448	-0.1662136	-0.2842765	-0.2874677
##	2604	0.0031818	-0.0632556	-0.0197814	-0.0619719
##	2605	-0.2201876	0.1102711	-0.2044282	0.0731433
##	2606	0.2515935	-0.1179361	0.2289698	-0.0248696
##	2607	-0.0695190	-0.1105951	-0.1297972	-0.1447809
##	2608	-0.2634320	0.0941096	-0.2106061	-0.0929258
##	2609	0.3895000	-0.5918097	-0.0568185	-0.2671643
##	2610	-0.1145527	-0.1482999	-0.1667510	-0.1857296
##	2611	0.0494061	0.0201147	0.0556626	0.0396443
##	2612	-0.1886105	0.2174745	-0.1372012	0.1458642
##	2613	-0.0282165	0.0217656	-0.0216801	0.0130057
##	2614	0.1312909	-0.0621041	0.1071407	-0.0195994
##	2700	0.1466248	-0.0803265	0.1186238	-0.0243808
##	2701	0.0873261	-0.0179826	0.0844261	-0.0036385
##	2702	-0.1122035	-0.0285811	-0.1190100	-0.0587086
##	2703	0.1177124	0.0536749	0.1218062	0.0635914
##	2704	0.0086590	-0.1581052	0.0013193	-0.1581486
##	2705	0.1111236	0.0958791	0.1234262	0.1195005
##	2706	0.0652880	0.0728218	0.0779524	0.0860446
##	2707	0.1449938	-0.1716050	0.0762198	-0.1125358
##	2708	0.0144165	-0.0048173	0.0138828	-0.0030889
##	2709	NA	NA	NA	NA
##	2710	-0.0707933	0.0054286	-0.0693875	-0.0129082
##	2711	0.1205911	0.0864835	0.1306632	0.1019593
##	2712	0.0244239	-0.0461571	0.0221264	-0.0444134
##	2713	0.0821744	-0.0127288	0.0805786	-0.0022624
##	2714	0.1131050	-0.0100101	0.1087701	0.0376355
##	2715	0.0162862	0.0463040	0.0195948	0.0481061
##	2716	-0.0707736	-0.0362602	-0.0753187	-0.0463414
##	2717	-0.0377843	-0.0007492	-0.0379718	-0.0098559
##	2718	0.1951563	0.0103952	0.1988527	0.0839452
##	2719	0.0726869	-0.0251549	0.0631042	0.0014316
##	2720	-0.0322420	-0.0997671	-0.0357234	-0.1013347
##	2721	0.0068665	-0.0035634	0.0066543	-0.0031161
##	2722	-0.0316602	-0.0195577	-0.0321871	-0.0203485
##	2723	-0.0446017	-0.0425053	-0.0462157	-0.0448690
##	2724	0.0605768	-0.0513632	0.0595412	-0.0497058
##	2725	-0.0081193	0.0100598	-0.0071168	0.0089911
##	2726	0.0362987	-0.0113576	0.0362571	-0.0110959
##	2727	-0.0205002	-0.0086268	-0.0210037	-0.0102512
##	2728	0.0868759	0.0526698	0.0973779	0.0723935
##	2729	0.1394094	0.0820781	0.1646262	0.1231358

## 2730	0.0141676	-0.0117540	0.0123633	-0.0092449
## 2731	0.0376691	-0.0444290	0.0315627	-0.0376648
## 2732	0.0119853	0.0191946	0.0166415	0.0229588
## 2733	0.0662053	0.0172759	0.0739592	0.0490437
## 2734	0.0647391	-0.0255189	0.0589698	-0.0081300
## 2735	0.0984268	0.0823736	0.1182880	0.1059499
## 2736	0.0580267	-0.0830621	0.0418917	-0.0704009
## 2737	0.0590019	-0.0093461	0.0580785	-0.0005504
## 2738	0.1036756	0.0305411	0.1081321	0.0460549
## 2739	0.0660404	-0.0080566	0.0638843	0.0091060
## 2740	0.1226477	0.0429591	0.1311442	0.0775522
## 2741	-0.0861067	-0.1294754	-0.1099237	-0.1508784
## 2742	0.0844691	-0.0233870	0.0784407	-0.0044020
## 2743	0.0471492	-0.0418376	0.0372916	-0.0308969
## 2744	NA	NA	NA	NA
## 2745	0.0601945	0.0570866	0.0665316	0.0650097
## 2746	0.1004764	0.0869043	0.1105220	0.1049256
## 2747	-0.0724819	-0.1920159	-0.0792859	-0.1974534
## 2748	-0.0063762	-0.0004865	-0.0063746	-0.0024409
## 2800	0.0277848	-0.0246027	0.0240802	-0.0190159
## 2801	0.0553916	-0.0535597	0.0501092	-0.0492656
## 2802	-0.0022960	-0.0358916	-0.0047977	-0.0360875
## 2803	0.0508009	0.1034671	0.0568714	0.1070069
## 2804	-0.0027288	-0.0409018	-0.0064885	-0.0412458
## 2805	-0.0064039	-0.0500354	-0.0153589	-0.0513464
## 2806	0.0935290	0.0837691	0.1078355	0.0999521
## 2807	0.0250777	-0.0389830	0.0230283	-0.0409871
## 2808	0.0980751	0.0176876	0.1016783	0.0418537
## 2809	-0.0050648	-0.0106350	-0.0064266	-0.0114170
## 2900	0.0495814	0.1069085	0.0958257	0.1252026
## 2901	NA	NA	-0.2208749	-0.1857275
## 2902	-0.0737227	-0.2159781	-0.1313387	-0.2327081
## 2903	NA	NA	NA	NA
## 2904	NA	NA	NA	NA
## 2905	0.0057421	0.0689657	0.0246683	0.0703113
## 2906	0.2624616	-0.4886304	0.0618454	-0.4083575
## 2907	0.2344304	-0.0830846	0.1971081	0.0173024
## 2908	0.2042242	-0.9321186	0.0479816	-0.9278685
## 2909	0.0615402	0.0391210	0.0707732	0.0525548
## 2910	0.1265307	-0.1028753	0.0464040	-0.0085190
## 2911	-0.0063161	-0.0800091	-0.0487079	-0.0832364
## 2912	0.3098843	-0.0428614	0.2952296	-0.0125910
## 2913	0.1166954	0.1475156	0.1176230	0.1472550
## 2914	-0.1213128	0.1808948	0.0047449	0.1052166
## 2915	NA	NA	NA	NA
## 2916	0.0225122	-0.0868911	0.0044225	-0.0828747
## 2917	0.4242321	0.0978917	0.4296580	0.1194242
## 2918	NA	NA	NA	NA
## 2919	0.0073792	-0.0919839	-0.0294266	-0.0898106
## 2920	NA	NA	NA	NA

##	2921	-0.0158219	0.1247925	0.0071356	0.1218983
##	2922	0.2400526	0.0851425	0.2851423	0.1866185
##	2923	0.2886385	-0.2571695	0.2277194	-0.2146900
##	3000	0.0301666	-0.2562813	-0.0549650	-0.2459545
##	3001	NA	NA	NA	NA
##	3002	0.0187429	0.0400983	0.0184988	0.0399045
##	3003	0.0085235	-0.1324836	-0.0135729	-0.1305594
##	3004	0.0236433	-0.0421217	0.0212630	-0.0402592
##	3005	-0.0189102	0.0508233	-0.0154524	0.0492397
##	3100	0.0225067	0.0631463	0.0335231	0.0687325
##	3101	0.0371231	0.0602080	0.0490860	0.0672179
##	3102	-0.0388205	0.1888155	0.0159901	0.1752541
##	3103	-0.0323389	0.0348532	-0.0221257	0.0227997
##	3104	0.0635161	-0.0196117	0.0616140	-0.0103855
##	3105	-0.0303260	0.0041030	-0.0299280	0.0017065
##	3106	-0.0027682	0.0576724	0.0089496	0.0569581
##	3107	0.0099440	0.0323058	0.0110731	0.0326682
##	3108	-0.0663165	0.0091380	-0.0646251	-0.0013978
##	3109	-0.1675800	-0.0765361	-0.1870390	-0.1280694
##	3110	0.0434065	-0.0077722	0.0429626	-0.0038548
##	3200	0.0197847	0.0813710	0.0433113	0.0871376
##	3201	0.1522082	-0.0295587	0.1372096	0.0499967
##	3202	0.0625124	0.0212528	0.0696997	0.0425377
##	3203	0.0243580	-0.0052638	0.0235893	-0.0018834
##	3204	0.0188030	-0.0557542	-0.0029483	-0.0491360
##	3205	0.0561566	0.0186082	0.0601538	0.0309252
##	3206	0.0568829	0.0556684	0.0660096	0.0650071
##	3207	0.0869268	-0.0541484	0.0640068	-0.0185857
##	3300	-0.0328611	-0.0195078	-0.0463910	-0.0424964
##	3301	-0.0221743	0.0912529	0.0381059	0.0769676
##	3302	-0.0570425	-0.0812008	-0.1045911	-0.1174045
##	3303	0.1079412	-0.0425992	0.0775591	0.0362165
##	3304	0.0267134	0.0013700	0.0274885	0.0162033
##	3305	0.0628606	0.0041177	0.0654892	0.0449846
##	3306	0.0978143	-0.0181226	0.0903847	0.0195622
##	3307	0.1516832	0.0766784	0.1744708	0.1213860
##	3308	0.0594455	-0.0541595	0.0215106	-0.0121601
##	3309	0.0190517	-0.0358073	-0.0071590	-0.0219217
##	3310	0.1464610	-0.0669748	0.0980749	0.0367313
##	3311	0.0094229	0.1789261	0.1028450	0.1840437
##	3312	0.0210536	0.0256181	0.0398629	0.0412275
##	3313	0.0722579	0.0486744	0.0911069	0.0793564
##	3314	0.0046801	-0.0003932	0.0043654	0.0032918
##	3315	0.1176578	-0.0716130	0.0621846	0.0190193
##	3316	-0.0038810	0.0491491	0.0374035	0.0457889
##	3317	0.0982418	0.0614719	0.1406360	0.1275525
##	3318	0.1123731	-0.1298362	0.0107328	-0.0524928
##	3319	-0.0273734	0.0809264	0.0089779	0.0689754
##	3320	0.0925754	-0.0065923	0.0869696	0.0732253
##	3321	-0.0351826	0.1326261	0.0366614	0.1126598

```

## 3322 0.0726815 -0.1119731 0.0062282 -0.0675848
## 3400 0.0667507 0.0599846 0.0701875 0.0649535
## 3401 NA NA NA NA
## 3402 NA NA NA NA
## 3403 0.1654920 0.1665628 0.1675094 0.1699775
## 3404 -0.0501148 0.0940602 -0.0434232 0.0884334
## 3500 0.0847377 -0.0634025 0.0583087 -0.0077589
## 3501 NA NA NA NA
## 3502 NA NA NA NA
## 3503 NA NA NA NA
## 3504 -0.0435600 0.1722313 0.0023573 0.1470944
## 3505 NA NA NA NA
## 3506 NA NA NA NA
## 3600 0.1802795 -0.0854085 0.1358857 0.0067294
## 3601 0.1449653 -0.0419784 0.1212293 0.0375598
## 3602 -0.1223529 -0.4102331 -0.3124114 -0.4454049
## 3603 0.2691303 -0.1114099 0.2525496 -0.0618250
## 3604 0.0743210 0.0920483 0.1210392 0.1269027
## 3605 -0.0634502 0.1700155 -0.0081121 0.1405093
## 3606 NA NA NA NA
## 3607 -0.0712139 -0.2314920 -0.1541509 -0.2556917
## 3608 NA NA NA NA
## 3609 0.2011092 -0.0138407 0.1967164 0.0411864
## 3610 0.1969539 0.0150448 0.2018078 0.0825968
## 3611 0.1404133 0.0818758 0.1650030 0.1220564
## 3612 -0.0156245 -0.0331761 -0.0270881 -0.0389686
## 3613 0.0628534 0.0345792 0.0666527 0.0405359
## 3614 -0.0402419 -0.0169673 -0.0432010 -0.0273988
## 3615 NA NA NA NA
## 3616 0.0497218 -0.0726777 0.0313261 -0.0647777

```

```
print(RegP)
```

```

##          FFA1p      FLA1p          2p          3p          4p          5+p
## 1000 7.379e-01 2.539e-01 4.296e-13 6.849e-30 5.729e-46 1.537e-78
## 1100 4.595e-01 9.868e-03 1.506e-16 2.252e-32 7.617e-32 3.970e-47
## 1101 1.727e-01 1.268e-01 3.116e-04 2.009e-04 3.901e-05 9.279e-06
## 1102 8.078e-01 5.145e-01 1.106e-06 4.658e-11 2.424e-14 7.777e-26
## 1103 2.324e-01 5.514e-01 2.769e-09 3.460e-14 1.584e-12 8.979e-17
## 1104 1.058e-01 7.412e-01 3.311e-04 4.601e-04 4.217e-08 1.369e-14
## 1105 2.580e-01 2.099e-03 4.074e-33 1.449e-49 7.428e-64 5.575e-77
## 1106 7.194e-01 7.551e-01 3.838e-14 9.087e-20 4.082e-31 9.606e-32
## 1107 4.390e-01 7.175e-01 2.012e-01 2.720e-04 6.244e-06 2.317e-05
## 1108 6.046e-02 9.280e-01 5.766e-02 2.940e-01 1.129e-02 6.147e-04
## 1109 9.835e-01 2.581e-01 8.883e-13 9.730e-17 6.828e-20 6.571e-33
## 1110 6.280e-03 4.617e-05 5.891e-14 3.546e-30 4.543e-46 3.212e-70
## 1111 4.107e-01 2.141e-01 9.472e-01 8.192e-02 2.199e-01 1.982e-01
## 1200 6.701e-01 1.004e-01 2.570e-03 2.224e-02 4.511e-02 6.136e-04
## 1201 7.907e-01 6.958e-01 2.901e-15 4.566e-12 9.385e-09 7.199e-11
## 1202 4.388e-01 4.024e-01 1.057e-05 3.004e-03 1.554e-04 6.823e-01

```

##	1203	3.661e-01	5.756e-01	2.056e-10	2.883e-18	1.145e-16	2.249e-19
##	1204	5.046e-01	7.694e-01	2.285e-04	7.115e-07	5.301e-04	6.087e-02
##	1205	9.911e-01	3.996e-01	2.697e-01	4.472e-02	2.447e-01	1.059e-01
##	1206	2.816e-01	4.329e-01	8.341e-01	4.310e-03	9.843e-01	3.550e-01
##	1207	9.486e-02	2.565e-01	1.107e-08	4.739e-10	7.577e-20	3.196e-18
##	1208	6.408e-01	5.621e-01	6.445e-02	5.900e-01	1.601e-15	2.951e-10
##	1209	2.445e-01	6.918e-01	9.601e-01	4.545e-05	1.063e-04	8.698e-01
##	1210	3.199e-01	6.935e-01	6.308e-07	1.206e-05	1.663e-03	2.964e-04
##	1211	8.934e-01	5.389e-01	2.896e-02	3.651e-06	3.732e-03	9.813e-02
##	1212	4.082e-01	2.133e-01	4.095e-02	4.420e-05	2.667e-01	2.929e-01
##	1213	8.871e-01	7.777e-02	2.163e-03	1.870e-03	5.632e-05	6.269e-01
##	1300	3.589e-01	2.551e-02	4.935e-08	3.050e-05	4.795e-08	1.607e-13
##	1301	NA	NA	NA	NA	NA	NA
##	1302	9.703e-01	8.857e-01	8.103e-04	2.276e-03	1.514e-01	2.917e-03
##	1303	2.353e-01	7.521e-01	9.745e-07	2.695e-08	7.166e-10	1.840e-17
##	1304	7.001e-01	8.056e-01	3.021e-04	2.121e-06	1.839e-07	2.104e-07
##	1305	7.165e-02	9.344e-01	8.035e-13	1.472e-15	9.591e-19	6.749e-24
##	1306	2.377e-01	2.874e-01	6.878e-08	5.970e-13	4.583e-13	1.014e-17
##	1307	1.521e-01	1.382e-03	1.528e-02	1.179e-04	3.680e-06	1.419e-11
##	1308	4.637e-01	3.260e-02	4.607e-04	4.297e-06	7.368e-06	1.485e-10
##	1309	2.965e-01	6.399e-02	3.035e-01	1.471e-01	3.448e-01	5.798e-04
##	1310	6.330e-01	6.603e-01	2.410e-01	2.616e-02	9.918e-03	9.838e-07
##	1311	2.794e-04	1.773e-03	3.540e-03	3.221e-05	1.022e-08	1.452e-14
##	1312	8.590e-02	3.198e-04	3.366e-04	2.715e-06	4.712e-08	5.593e-16
##	1313	8.944e-01	3.172e-01	2.397e-02	2.375e-04	1.235e-04	6.272e-07
##	1314	2.490e-01	1.381e-03	2.710e-01	7.230e-02	6.324e-04	4.169e-06
##	1315	6.949e-01	8.504e-01	5.316e-03	3.232e-03	3.669e-02	1.682e-04
##	1400	6.686e-01	7.421e-01	3.904e-05	7.919e-13	1.536e-03	3.299e-02
##	1401	7.073e-01	4.040e-01	2.014e-03	2.056e-01	1.086e-01	3.284e-01
##	1402	5.558e-01	9.930e-03	1.846e-04	7.747e-07	2.104e-01	8.301e-01
##	1403	1.560e-01	6.575e-01	1.656e-04	2.869e-05	9.671e-03	3.488e-02
##	1404	6.144e-02	7.052e-01	4.397e-02	1.301e-01	9.221e-04	6.687e-04
##	1405	3.783e-01	5.920e-01	3.225e-04	2.267e-04	1.716e-04	5.522e-03
##	1406	3.280e-01	1.754e-01	2.417e-09	6.031e-07	2.452e-01	1.279e-01
##	1407	1.100e-01	7.391e-01	6.723e-10	3.675e-09	2.058e-11	2.223e-02
##	1408	6.705e-03	1.911e-01	1.670e-03	9.939e-09	8.936e-03	8.296e-01
##	1409	6.285e-02	1.244e-01	2.286e-01	3.152e-02	5.144e-01	4.533e-01
##	1410	1.336e-02	4.491e-01	2.256e-02	5.537e-06	4.033e-02	4.303e-01
##	1500	4.505e-01	9.959e-02	1.008e-09	2.824e-07	6.592e-07	3.866e-08
##	1501	5.622e-01	7.772e-09	3.493e-01	5.508e-01	6.242e-03	7.772e-03
##	1502	6.974e-01	2.088e-01	2.457e-06	4.102e-09	2.506e-10	4.672e-10
##	1503	3.578e-03	1.929e-01	2.258e-01	8.396e-01	7.728e-01	3.122e-01
##	1504	6.029e-01	1.020e-01	7.181e-01	6.047e-01	7.008e-01	8.899e-02
##	1505	1.886e-01	3.360e-01	5.086e-01	7.250e-01	7.329e-01	5.439e-01
##	1506	3.743e-01	1.166e-01	3.029e-03	1.747e-03	1.853e-02	1.805e-05
##	1507	8.350e-01	9.358e-01	6.853e-02	6.072e-02	5.807e-01	9.681e-01
##	1508	1.550e-01	7.063e-01	1.604e-01	8.587e-04	9.172e-02	2.125e-02
##	1600	1.773e-01	2.065e-01	5.963e-04	1.157e-01	9.396e-04	4.020e-09
##	1601	NA	NA	NA	NA	NA	NA
##	1602	4.822e-01	1.584e-01	6.716e-01	1.488e-01	7.300e-02	2.181e-02

## 1603	1.192e-01	6.966e-01	9.173e-01	6.388e-01	9.636e-02	2.535e-02
## 1604	2.920e-01	2.318e-01	5.301e-02	7.701e-02	1.173e-01	4.134e-01
## 1605	2.724e-01	1.568e-01	1.876e-01	2.682e-02	1.241e-02	1.406e-03
## 1606	2.751e-02	2.115e-03	2.511e-02	1.014e-03	4.285e-03	2.998e-05
## 1607	5.349e-02	2.478e-01	7.710e-01	4.836e-01	4.074e-01	1.246e-01
## 1700	3.378e-03	5.762e-01	9.279e-02	4.821e-04	1.105e-02	1.811e-01
## 1701	7.589e-02	8.948e-02	6.011e-01	9.026e-01	1.867e-04	1.310e-01
## 1702	8.052e-01	1.016e-01	6.210e-02	2.514e-03	9.979e-02	3.706e-02
## 1703	4.041e-01	5.959e-01	1.710e-02	1.466e-05	1.288e-04	2.503e-06
## 1704	8.206e-01	2.010e-01	1.100e-06	6.527e-06	2.393e-03	1.549e-01
## 1705	6.406e-02	6.367e-01	8.986e-01	9.272e-06	1.104e-03	1.379e-02
## 1706	2.643e-01	1.267e-01	1.869e-11	8.671e-16	6.870e-12	1.320e-09
## 1707	3.147e-01	6.387e-01	5.686e-01	9.881e-02	8.432e-01	1.242e-01
## 1708	9.696e-01	5.820e-01	2.551e-01	5.601e-04	1.336e-03	5.155e-05
## 1709	2.402e-01	4.019e-01	1.719e-03	8.790e-07	7.701e-04	3.619e-03
## 1710	8.788e-01	7.918e-01	6.164e-06	5.938e-09	8.065e-05	2.863e-02
## 1711	9.722e-01	8.285e-01	5.975e-01	8.462e-01	4.947e-02	5.995e-01
## 1712	6.569e-01	8.155e-01	4.985e-05	1.543e-11	8.731e-08	1.421e-06
## 1800	4.445e-01	3.289e-01	2.427e-02	4.413e-02	3.571e-01	3.621e-01
## 1801	NA	NA	NA	NA	NA	NA
## 1802	7.002e-01	3.110e-01	5.612e-02	7.103e-04	2.721e-01	1.138e-01
## 1803	3.975e-01	7.732e-01	3.078e-01	5.394e-01	3.714e-02	3.826e-01
## 1804	3.439e-01	9.249e-01	2.557e-06	2.802e-04	5.580e-03	2.421e-01
## 1900	5.925e-02	3.609e-01	1.259e-04	3.046e-06	4.857e-02	3.046e-04
## 1901	2.457e-01	5.548e-02	1.029e-01	6.301e-04	1.044e-03	6.820e-03
## 1902	1.864e-01	2.208e-01	1.163e-04	9.007e-08	5.444e-05	5.640e-10
## 1903	3.442e-01	4.720e-01	3.172e-01	6.278e-03	6.481e-01	1.275e-02
## 1904	7.558e-01	9.309e-01	1.154e-02	2.598e-04	2.172e-02	1.144e-02
## 1905	3.218e-01	3.654e-01	2.219e-01	3.303e-02	7.237e-01	4.357e-02
## 1906	1.828e-02	6.785e-01	3.127e-03	1.703e-05	3.560e-04	1.248e-07
## 1907	1.456e-01	6.140e-01	1.480e-05	5.013e-12	9.676e-11	3.825e-14
## 1908	7.065e-03	2.511e-01	1.316e-02	1.144e-08	4.862e-09	1.459e-08
## 1909	6.752e-01	2.849e-01	1.386e-05	7.120e-21	4.342e-06	4.896e-04
## 1910	2.455e-01	4.193e-01	7.796e-05	6.467e-05	1.583e-05	1.361e-10
## 1911	3.240e-02	7.600e-03	2.341e-02	4.290e-05	7.018e-05	5.773e-03
## 1912	1.393e-02	8.410e-01	1.036e-05	9.998e-08	1.134e-06	3.090e-03
## 1913	2.989e-01	6.925e-01	5.973e-02	4.017e-02	8.747e-02	1.627e-01
## 2000	9.626e-01	3.777e-01	6.506e-02	1.579e-01	2.314e-02	1.096e-01
## 2001	6.167e-01	5.357e-03	9.828e-03	1.291e-05	2.968e-02	1.808e-04
## 2002	7.892e-01	1.332e-02	1.812e-11	6.834e-13	2.204e-05	4.512e-05
## 2003	8.068e-01	5.157e-03	1.364e-03	3.160e-04	6.649e-01	6.429e-01
## 2100	NA	NA	NA	NA	NA	NA
## 2101	NA	NA	NA	NA	NA	NA
## 2102	5.284e-01	4.748e-01	1.196e-06	5.591e-10	1.982e-09	2.081e-03
## 2103	7.965e-01	4.851e-01	1.530e-04	1.101e-06	3.719e-05	1.433e-02
## 2104	8.467e-01	1.260e-01	3.083e-01	3.639e-03	7.881e-02	8.986e-04
## 2105	4.477e-01	1.015e-01	4.667e-03	6.252e-04	3.094e-05	1.617e-03
## 2200	NA	NA	NA	NA	NA	NA
## 2201	5.068e-01	1.685e-01	7.562e-02	2.957e-06	2.215e-02	1.420e-13
## 2202	3.567e-01	9.585e-01	6.315e-04	1.140e-07	1.695e-02	3.865e-02

##	2203	8.650e-01	6.013e-01	1.117e-01	1.563e-03	7.443e-03	9.574e-01
##	2204	6.869e-01	8.979e-01	1.032e-01	3.350e-02	7.488e-04	1.665e-02
##	2205	8.975e-01	5.348e-01	6.654e-04	6.408e-05	2.925e-03	1.744e-05
##	2206	2.665e-02	2.934e-04	1.474e-01	1.761e-01	5.200e-01	5.851e-02
##	2207	9.666e-01	9.356e-02	7.019e-04	9.311e-07	7.157e-03	4.599e-01
##	2208	7.388e-01	8.489e-01	5.762e-03	2.544e-02	5.667e-02	2.249e-01
##	2209	1.045e-01	5.017e-01	5.293e-02	2.705e-03	1.694e-06	2.100e-04
##	2210	4.696e-02	9.361e-02	6.461e-05	2.036e-11	3.534e-08	1.235e-12
##	2211	2.375e-01	6.822e-01	9.880e-04	6.345e-06	3.297e-05	9.338e-09
##	2212	5.861e-01	5.459e-01	7.433e-01	2.845e-01	9.884e-01	1.501e-01
##	2213	7.115e-02	1.028e-01	2.629e-01	8.107e-05	1.191e-03	1.160e-03
##	2214	6.653e-01	7.909e-01	1.958e-02	4.096e-02	9.382e-02	3.125e-02
##	2215	1.500e-01	4.173e-01	2.846e-04	1.000e-03	8.892e-02	1.715e-05
##	2216	2.032e-02	8.610e-02	5.742e-01	1.129e-01	2.116e-01	2.955e-01
##	2300	5.629e-01	4.590e-02	9.350e-08	4.397e-11	1.240e-14	1.714e-21
##	2301	8.465e-01	9.044e-01	2.400e-02	1.404e-01	1.277e-01	9.979e-01
##	2302	4.535e-02	3.419e-02	1.423e-01	2.867e-03	4.701e-05	8.641e-06
##	2303	2.154e-01	6.939e-01	7.167e-16	5.777e-18	3.135e-22	5.161e-33
##	2304	4.841e-01	1.812e-03	6.185e-07	8.357e-05	6.039e-04	4.060e-10
##	2305	1.950e-02	6.418e-01	5.360e-06	3.492e-07	2.890e-05	8.461e-08
##	2306	6.728e-01	1.347e-02	1.153e-02	1.048e-02	4.403e-04	4.207e-04
##	2307	3.206e-01	3.167e-01	2.171e-01	7.336e-01	3.760e-02	5.892e-03
##	2308	6.286e-02	5.045e-01	3.073e-07	2.896e-11	1.426e-09	1.381e-13
##	2309	9.514e-02	1.427e-01	1.581e-04	3.310e-10	2.964e-15	2.589e-12
##	2310	4.188e-01	1.816e-01	3.543e-04	4.755e-06	2.813e-05	3.603e-09
##	2311	4.485e-02	1.836e-01	3.248e-04	6.328e-05	1.918e-07	9.530e-07
##	2312	7.920e-01	8.489e-01	2.364e-12	1.173e-20	1.997e-11	4.491e-15
##	2400	4.718e-02	4.775e-01	1.433e-01	6.777e-02	6.636e-02	2.022e-01
##	2401	NA	NA	NA	NA	NA	NA
##	2402	7.737e-01	4.586e-01	3.624e-15	6.651e-18	1.975e-17	8.883e-25
##	2403	5.840e-03	2.048e-02	1.071e-01	1.727e-02	2.110e-03	1.832e-06
##	2404	1.480e-01	2.065e-01	1.555e-02	1.153e-04	2.489e-05	9.309e-07
##	2405	8.313e-02	7.250e-01	1.372e-02	1.555e-03	2.466e-05	1.035e-09
##	2406	1.635e-01	8.687e-01	5.916e-01	7.429e-01	3.536e-01	2.442e-01
##	2500	2.163e-04	5.274e-01	8.768e-07	4.835e-02	2.308e-05	2.256e-15
##	2501	9.600e-01	4.955e-01	5.159e-03	6.423e-03	9.700e-03	1.224e-02
##	2502	8.276e-01	5.129e-01	1.035e-01	3.707e-02	5.121e-04	4.195e-04
##	2503	2.201e-01	3.615e-01	8.025e-05	2.833e-06	1.131e-07	3.202e-08
##	2504	4.237e-02	8.622e-01	7.831e-01	5.186e-01	8.031e-02	3.671e-02
##	2505	2.326e-01	5.662e-01	3.189e-01	2.554e-01	2.777e-01	9.603e-02
##	2506	6.084e-01	1.283e-01	2.212e-02	5.401e-03	2.114e-03	1.163e-02
##	2507	6.710e-01	6.334e-01	1.224e-09	1.722e-13	1.706e-14	3.357e-16
##	2508	1.143e-01	3.417e-01	7.970e-01	2.655e-01	1.382e-01	8.044e-02
##	2600	6.674e-01	5.387e-01	1.184e-08	1.917e-03	1.996e-01	1.465e-02
##	2601	5.922e-01	3.571e-01	3.974e-01	3.394e-01	2.222e-01	9.473e-01
##	2602	2.792e-01	6.615e-01	4.627e-03	8.923e-01	6.592e-01	9.775e-01
##	2603	1.608e-01	5.309e-01	4.765e-03	5.194e-01	4.501e-01	8.896e-01
##	2604	7.324e-01	2.069e-01	8.295e-08	4.365e-05	5.206e-05	2.968e-02
##	2605	2.591e-03	2.127e-01	2.016e-01	2.676e-01	1.067e-01	5.812e-04
##	2606	3.318e-02	4.599e-01	2.728e-01	5.958e-01	1.462e-01	3.757e-01

##	2607	5.611e-01	1.490e-01	3.881e-01	1.404e-01	4.441e-01	7.917e-01
##	2608	1.576e-02	3.065e-01	1.004e-02	5.532e-01	1.344e-06	2.164e-01
##	2609	6.509e-02	2.683e-03	1.706e-02	1.549e-04	5.089e-01	5.446e-03
##	2610	1.752e-01	4.193e-01	8.989e-01	8.199e-04	1.333e-01	7.308e-08
##	2611	9.391e-01	7.742e-01	9.040e-05	1.413e-05	5.439e-04	4.938e-04
##	2612	2.024e-01	3.194e-01	6.154e-02	8.061e-02	6.448e-02	1.414e-02
##	2613	1.660e-01	5.997e-01	3.065e-07	1.700e-13	1.413e-09	2.216e-04
##	2614	7.924e-02	2.479e-01	4.585e-01	4.175e-04	1.037e-01	1.475e-02
##	2700	2.161e-01	3.057e-01	3.709e-72	2.360e-165	1.473e-208	0.000e+00
##	2701	5.361e-03	8.894e-01	8.192e-07	1.567e-13	2.412e-15	3.199e-27
##	2702	8.158e-02	9.266e-01	1.554e-01	2.414e-02	3.449e-03	5.157e-06
##	2703	2.100e-03	1.175e-01	3.171e-07	1.693e-07	1.919e-10	1.884e-17
##	2704	9.226e-01	5.313e-01	1.058e-03	6.176e-12	5.263e-09	4.375e-25
##	2705	1.605e-04	3.865e-06	4.676e-04	2.070e-09	2.549e-15	2.025e-41
##	2706	2.406e-01	7.823e-03	3.370e-07	1.227e-14	1.079e-15	3.451e-46
##	2707	2.257e-02	2.976e-02	1.448e-10	3.628e-13	1.911e-14	1.706e-22
##	2708	7.929e-01	8.173e-01	1.538e-02	1.047e-04	2.357e-06	3.606e-09
##	2709	NA	NA	NA	NA	NA	NA
##	2710	3.364e-01	9.111e-01	8.001e-01	7.612e-01	8.872e-01	5.966e-01
##	2711	2.586e-02	5.532e-02	7.569e-05	8.762e-10	9.987e-12	9.820e-20
##	2712	6.337e-01	1.976e-01	2.304e-01	2.903e-03	1.763e-03	4.980e-07
##	2713	4.023e-02	6.660e-01	1.845e-02	1.169e-04	3.324e-04	5.648e-05
##	2714	2.634e-01	7.554e-01	3.959e-07	2.384e-05	2.324e-14	2.324e-20
##	2715	8.607e-01	2.847e-01	1.622e-02	1.878e-03	1.069e-03	1.503e-06
##	2716	3.865e-03	1.698e-01	7.337e-01	1.706e-01	3.661e-02	1.084e-02
##	2717	1.277e-01	4.992e-01	1.385e-04	2.440e-05	1.037e-06	1.030e-11
##	2718	6.637e-03	7.630e-01	8.471e-02	1.599e-02	7.604e-03	1.300e-04
##	2719	6.107e-01	6.687e-01	3.848e-13	3.512e-17	1.845e-24	2.836e-38
##	2720	1.005e-01	7.353e-02	3.138e-03	1.542e-06	4.258e-05	5.981e-13
##	2721	4.254e-01	4.431e-01	4.577e-02	6.748e-02	1.383e-02	3.356e-04
##	2722	7.318e-01	8.395e-01	9.120e-01	6.129e-01	9.036e-02	6.227e-05
##	2723	2.447e-03	1.588e-01	2.370e-08	2.368e-08	1.384e-11	6.145e-16
##	2724	9.363e-01	4.946e-02	1.283e-02	5.350e-08	9.599e-07	6.455e-16
##	2725	5.621e-01	3.456e-01	3.496e-02	9.932e-03	1.884e-05	2.715e-08
##	2726	5.932e-01	8.351e-01	9.778e-15	3.568e-16	2.777e-22	1.299e-27
##	2727	2.791e-01	3.732e-01	6.402e-02	3.845e-03	3.476e-03	1.278e-08
##	2728	1.091e-04	5.493e-04	1.974e-06	1.261e-10	4.242e-13	1.134e-24
##	2729	6.726e-07	3.630e-05	2.061e-05	4.493e-06	1.697e-06	3.379e-12
##	2730	9.586e-01	5.732e-01	1.694e-08	9.656e-17	9.824e-17	2.349e-25
##	2731	2.488e-01	1.562e-01	3.145e-08	4.190e-12	2.343e-17	3.789e-26
##	2732	9.374e-01	2.302e-01	3.701e-02	1.757e-04	2.302e-07	7.918e-11
##	2733	3.441e-01	1.857e-01	9.926e-03	1.731e-03	3.088e-07	6.359e-08
##	2734	2.580e-01	9.200e-01	7.788e-14	1.188e-17	1.795e-14	3.949e-29
##	2735	3.377e-04	1.021e-05	3.138e-07	3.267e-09	3.445e-12	3.167e-21
##	2736	4.606e-01	3.830e-01	2.663e-16	5.015e-31	2.840e-46	5.359e-68
##	2737	1.034e-02	7.817e-01	4.602e-04	1.826e-04	1.622e-06	2.537e-12
##	2738	6.980e-07	3.551e-02	3.571e-63	7.436e-89	6.037e-95	1.740e-134
##	2739	7.383e-03	5.778e-01	5.222e-14	5.837e-31	2.189e-39	1.058e-57
##	2740	1.118e-03	2.491e-02	3.933e-07	2.137e-08	8.385e-14	1.702e-31
##	2741	3.843e-03	8.572e-02	2.699e-04	1.134e-13	3.542e-14	2.653e-32

##	2742	1.871e-01	3.543e-01	7.180e-07	1.676e-14	2.132e-19	7.268e-23
##	2743	2.765e-01	3.466e-01	4.725e-03	1.338e-01	2.363e-01	1.823e-02
##	2744	NA	NA	NA	NA	NA	NA
##	2745	6.294e-01	1.982e-01	9.537e-01	8.297e-02	5.980e-03	2.082e-03
##	2746	2.338e-03	1.384e-03	3.609e-10	8.068e-14	8.407e-18	5.043e-37
##	2747	5.635e-02	1.190e-02	1.750e-04	1.969e-07	3.508e-08	5.049e-15
##	2748	8.463e-01	8.382e-01	1.721e-01	6.421e-03	1.167e-03	3.762e-05
##	2800	6.666e-01	3.029e-01	3.277e-07	3.310e-08	7.992e-11	6.887e-17
##	2801	4.598e-01	7.492e-01	7.293e-01	8.846e-01	8.932e-01	3.645e-01
##	2802	5.347e-01	2.798e-01	1.929e-02	1.700e-02	8.877e-03	1.408e-05
##	2803	6.417e-01	1.475e-03	1.480e-01	8.521e-03	1.279e-03	1.208e-04
##	2804	6.843e-01	1.384e-01	4.631e-01	7.337e-01	7.087e-01	3.247e-01
##	2805	2.242e-01	2.008e-01	1.890e-04	1.813e-05	2.593e-05	7.369e-08
##	2806	1.883e-01	1.285e-01	3.264e-02	7.516e-03	3.154e-02	1.598e-03
##	2807	9.357e-01	5.271e-01	1.735e-01	1.900e-01	3.145e-01	2.077e-02
##	2808	3.208e-04	1.459e-01	1.096e-03	5.346e-05	5.409e-07	1.244e-12
##	2809	2.860e-01	9.064e-01	8.370e-03	1.079e-02	1.767e-05	5.221e-09
##	2900	7.968e-01	8.498e-05	1.571e-12	2.745e-22	8.760e-25	6.327e-28
##	2901	2.432e-03	2.993e-01	1.689e-05	5.509e-03	4.813e-01	1.048e-03
##	2902	3.189e-01	1.522e-01	5.127e-02	6.271e-05	6.348e-07	1.608e-13
##	2903	NA	NA	NA	NA	NA	NA
##	2904	NA	NA	NA	NA	NA	NA
##	2905	8.616e-01	1.042e-01	7.294e-01	7.306e-02	1.150e-01	3.052e-02
##	2906	3.059e-01	5.886e-01	4.192e-14	7.170e-10	4.437e-11	1.272e-10
##	2907	4.532e-02	8.542e-01	1.454e-09	1.086e-10	1.550e-11	9.413e-19
##	2908	NA	NA	NA	NA	NA	NA
##	2909	1.962e-01	7.002e-01	6.772e-03	2.290e-03	6.097e-03	7.492e-05
##	2910	5.785e-01	7.460e-01	6.056e-02	6.458e-06	7.044e-04	1.895e-01
##	2911	9.399e-01	9.659e-01	3.990e-01	2.080e-04	2.744e-03	1.988e-03
##	2912	7.182e-02	7.197e-01	6.365e-02	3.216e-03	1.756e-04	6.594e-02
##	2913	3.470e-01	3.560e-03	6.302e-03	2.389e-04	4.887e-06	7.547e-09
##	2914	5.932e-01	1.698e-01	1.866e-02	6.335e-02	9.934e-01	7.860e-03
##	2915	NA	NA	NA	NA	NA	NA
##	2916	3.835e-01	1.223e-02	1.997e-08	1.220e-16	2.498e-14	1.488e-22
##	2917	2.809e-02	2.237e-01	9.122e-01	1.035e-02	4.707e-01	3.696e-02
##	2918	NA	NA	NA	NA	NA	NA
##	2919	8.338e-01	6.514e-01	4.968e-03	8.547e-02	2.391e-01	1.977e-02
##	2920	NA	NA	NA	NA	NA	NA
##	2921	7.944e-01	5.845e-02	6.338e-03	2.860e-02	5.209e-01	3.227e-01
##	2922	9.887e-02	7.920e-01	6.838e-01	9.117e-01	3.510e-01	5.704e-01
##	2923	2.398e-01	3.484e-01	9.601e-01	1.699e-01	6.349e-01	7.021e-01
##	3000	9.802e-01	3.967e-01	2.319e-02	2.027e-06	4.517e-07	1.872e-08
##	3001	NA	NA	NA	NA	NA	NA
##	3002	5.364e-01	4.008e-01	2.767e-01	5.252e-02	4.707e-02	6.847e-04
##	3003	9.947e-01	6.529e-02	1.357e-03	6.790e-12	3.518e-17	6.371e-16
##	3004	7.255e-01	7.709e-01	7.259e-21	6.893e-37	5.439e-39	1.519e-43
##	3005	2.048e-01	1.450e-01	5.700e-01	2.941e-02	5.792e-02	1.648e-02
##	3100	4.469e-01	6.098e-01	1.481e-07	1.546e-14	5.872e-11	3.410e-16
##	3101	6.902e-01	4.961e-01	2.212e-01	6.886e-02	7.509e-03	3.776e-03
##	3102	2.006e-01	3.321e-02	4.723e-04	8.998e-03	5.446e-05	2.104e-04

##	3103	1.280e-01	6.808e-01	3.290e-07	9.153e-07	6.065e-05	1.227e-02
##	3104	8.475e-02	5.380e-01	2.160e-02	3.090e-01	1.870e-01	6.354e-04
##	3105	4.613e-01	9.491e-01	5.411e-01	2.210e-01	2.766e-01	2.277e-01
##	3106	9.850e-01	5.278e-01	2.953e-01	1.862e-02	7.241e-01	1.199e-02
##	3107	6.143e-01	8.642e-01	5.453e-01	1.158e-01	5.365e-03	2.462e-07
##	3108	2.422e-01	7.599e-01	4.821e-02	4.150e-01	7.135e-01	2.548e-01
##	3109	2.758e-02	7.428e-01	5.441e-01	8.158e-04	3.244e-01	7.438e-01
##	3110	4.496e-01	6.993e-01	1.333e-01	1.090e-01	1.770e-02	6.505e-03
##	3200	7.295e-01	9.623e-04	5.814e-11	2.481e-10	2.903e-07	1.066e-13
##	3201	1.330e-02	6.796e-01	6.052e-05	6.081e-02	5.638e-04	5.055e-02
##	3202	9.520e-02	4.551e-01	3.954e-04	1.824e-06	6.514e-06	1.823e-03
##	3203	5.024e-01	9.730e-01	1.195e-15	1.396e-13	1.562e-20	2.160e-32
##	3204	4.921e-01	1.262e-02	3.620e-08	1.429e-10	1.097e-14	5.349e-21
##	3205	4.524e-02	3.863e-01	8.501e-05	4.581e-04	2.223e-07	1.366e-08
##	3206	5.173e-01	1.329e-01	2.342e-10	9.839e-08	1.277e-11	1.991e-10
##	3207	1.493e-02	4.062e-02	1.927e-13	1.363e-12	4.584e-08	1.614e-12
##	3300	2.571e-01	8.140e-01	2.824e-04	4.804e-04	7.935e-03	4.206e-02
##	3301	3.463e-01	1.625e-02	1.446e-03	2.849e-04	2.719e-01	2.256e-02
##	3302	1.851e-01	4.683e-01	2.219e-03	4.799e-06	1.040e-03	3.700e-02
##	3303	3.765e-02	5.219e-01	7.735e-09	9.732e-07	5.343e-01	7.404e-01
##	3304	7.665e-01	4.007e-01	6.140e-18	3.267e-28	6.305e-15	1.301e-15
##	3305	2.136e-02	4.974e-01	1.428e-11	3.642e-08	4.697e-03	3.295e-05
##	3306	2.230e-03	8.778e-01	2.779e-06	1.824e-08	4.247e-11	1.519e-12
##	3307	2.691e-02	6.560e-02	3.680e-04	4.181e-09	2.019e-08	6.980e-09
##	3308	4.328e-01	3.054e-01	5.271e-18	1.145e-26	5.475e-15	1.482e-16
##	3309	6.684e-01	9.775e-01	1.211e-15	1.715e-07	1.904e-12	1.358e-03
##	3310	6.986e-01	7.357e-01	2.241e-11	5.789e-21	1.851e-12	1.270e-26
##	3311	6.792e-01	3.426e-03	2.636e-06	2.253e-04	4.122e-03	1.452e-01
##	3312	6.742e-01	1.221e-01	5.349e-18	2.744e-23	6.529e-15	4.176e-05
##	3313	3.031e-01	3.903e-01	2.297e-02	1.639e-02	1.968e-01	3.222e-01
##	3314	8.555e-01	9.283e-01	1.254e-06	1.692e-06	1.602e-07	5.337e-01
##	3315	3.752e-02	3.781e-01	1.164e-08	2.547e-16	3.739e-03	1.331e-03
##	3316	4.915e-01	1.421e-01	1.187e-04	4.453e-04	1.647e-02	7.428e-03
##	3317	1.461e-01	3.775e-01	4.881e-02	8.155e-02	7.770e-03	2.735e-01
##	3318	6.519e-01	6.591e-01	2.138e-11	4.233e-05	5.131e-04	4.638e-04
##	3319	5.631e-01	1.011e-01	1.371e-03	4.801e-05	2.066e-02	7.632e-02
##	3320	2.407e-01	8.738e-01	5.828e-04	5.107e-01	2.940e-02	4.997e-02
##	3321	4.401e-01	7.635e-03	4.892e-01	1.293e-01	5.642e-01	1.425e-01
##	3322	2.494e-01	1.426e-01	1.612e-03	7.357e-03	5.875e-01	3.281e-01
##	3400	2.038e-01	6.562e-02	3.564e-03	2.615e-04	1.502e-04	2.148e-04
##	3401	NA	NA	NA	NA	NA	NA
##	3402	NA	NA	NA	NA	NA	NA
##	3403	2.268e-02	3.576e-02	4.725e-01	1.669e-01	2.869e-01	8.450e-01
##	3404	3.032e-01	4.356e-01	3.910e-03	1.702e-03	3.642e-02	2.264e-05
##	3500	5.157e-01	8.366e-01	7.200e-02	1.693e-05	1.975e-06	1.374e-04
##	3501	NA	NA	NA	NA	NA	NA
##	3502	NA	NA	NA	NA	NA	NA
##	3503	NA	NA	NA	NA	NA	NA
##	3504	7.336e-01	3.455e-01	5.882e-01	5.321e-01	8.264e-01	8.278e-01
##	3505	NA	NA	NA	NA	NA	NA

## 3506	NA	NA	NA	NA	NA	NA
## 3600	3.994e-02	3.929e-01	2.818e-02	3.629e-02	1.791e-03	1.717e-03
## 3601	4.045e-01	7.640e-01	6.106e-04	6.060e-04	9.606e-02	2.979e-01
## 3602	5.547e-01	1.831e-01	3.427e-02	5.436e-03	4.429e-10	3.921e-03
## 3603	1.909e-01	4.821e-01	1.157e-02	7.984e-02	2.805e-03	8.206e-04
## 3604	7.821e-01	4.015e-01	9.022e-03	7.010e-02	2.816e-01	4.749e-01
## 3605	3.299e-01	1.378e-01	1.967e-01	4.038e-01	1.163e-02	1.039e-01
## 3606	NA	NA	NA	NA	NA	NA
## 3607	6.677e-01	3.136e-01	2.617e-04	1.283e-03	1.544e-01	1.268e-04
## 3608	NA	NA	NA	NA	NA	NA
## 3609	6.816e-04	9.306e-01	5.664e-03	2.510e-02	4.302e-02	2.746e-02
## 3610	3.240e-01	9.600e-01	6.076e-02	3.239e-05	1.696e-06	2.887e-07
## 3611	2.451e-01	2.772e-01	3.755e-07	1.892e-08	2.445e-07	3.358e-11
## 3612	3.222e-02	6.114e-01	7.598e-07	2.712e-16	3.510e-22	1.410e-31
## 3613	5.729e-01	7.522e-01	7.188e-02	1.306e-01	7.725e-02	6.362e-02
## 3614	2.155e-01	9.317e-01	3.738e-01	6.844e-01	8.332e-01	6.946e-02
## 3615	NA	NA	NA	NA	NA	NA
## 3616	9.899e-01	1.976e-01	1.162e-03	7.406e-06	2.525e-08	5.871e-05
##	FFA2p	FLA2p	FFA3p	FLA4p		
## 1000	8.209e-07	2.022e-01	6.382e-06	6.059e-01		
## 1100	2.755e-03	1.435e-02	1.270e-02	5.523e-02		
## 1101	7.050e-01	1.137e-01	5.080e-01	9.821e-02		
## 1102	8.301e-02	2.913e-01	1.102e-01	4.069e-01		
## 1103	4.910e-03	2.016e-01	1.116e-02	5.649e-01		
## 1104	5.460e-02	7.877e-01	4.787e-02	5.660e-01		
## 1105	4.986e-01	1.559e-05	7.417e-01	2.211e-05		
## 1106	5.932e-02	6.443e-01	7.501e-02	9.048e-01		
## 1107	7.171e-01	5.222e-01	5.855e-01	4.405e-01		
## 1108	1.905e-01	7.709e-01	1.850e-01	6.825e-01		
## 1109	1.199e-01	1.774e-01	1.873e-01	2.662e-01		
## 1110	4.562e-01	5.581e-07	1.096e-01	2.710e-07		
## 1111	2.834e-01	1.789e-01	2.254e-01	1.364e-01		
## 1200	3.246e-01	5.685e-02	4.029e-01	4.595e-02		
## 1201	3.053e-01	9.917e-01	2.760e-01	6.121e-01		
## 1202	7.511e-01	7.679e-01	9.017e-01	9.825e-01		
## 1203	7.157e-05	2.918e-01	8.456e-06	1.161e-02		
## 1204	9.901e-01	3.184e-01	4.372e-01	NA		
## 1205	2.789e-01	5.478e-02	1.977e-01	1.418e-01		
## 1206	1.602e-01	2.869e-01	4.513e-01	9.886e-01		
## 1207	1.731e-02	1.534e-01	2.494e-07	1.601e-06		
## 1208	2.499e-01	9.788e-01	8.585e-04	1.855e-03		
## 1209	3.881e-02	2.154e-01	1.305e-01	6.232e-01		
## 1210	1.704e-02	2.493e-01	9.049e-03	2.630e-01		
## 1211	7.964e-01	4.932e-01	4.684e-01	3.287e-01		
## 1212	7.356e-01	3.461e-01	3.509e-01	2.151e-01		
## 1213	8.731e-01	1.633e-01	2.658e-02	1.180e-02		
## 1300	9.506e-01	5.260e-02	6.857e-01	4.881e-02		
## 1301	NA	NA	NA	NA		
## 1302	8.834e-01	8.194e-01	8.344e-01	7.879e-01		
## 1303	9.748e-01	8.702e-01	9.800e-01	8.710e-01		

```
## 1304 5.428e-01 6.473e-01 5.075e-01 6.018e-01
## 1305 3.727e-01 4.732e-01 3.274e-01 4.117e-01
## 1306 6.778e-01 1.394e-01 5.704e-01 1.292e-01
## 1307 6.039e-01 2.906e-04 4.532e-01 2.424e-04
## 1308 8.888e-01 8.358e-02 7.816e-01 8.305e-02
## 1309 3.447e-01 2.529e-02 1.974e-01 1.585e-02
## 1310 4.187e-01 4.154e-01 4.444e-01 4.411e-01
## 1311 1.643e-02 8.841e-04 6.258e-03 3.632e-04
## 1312 4.173e-01 1.559e-04 3.067e-01 1.280e-04
## 1313 9.502e-01 5.301e-01 9.295e-01 5.291e-01
## 1314 6.408e-01 2.277e-03 4.140e-01 1.796e-03
## 1315 7.709e-01 8.544e-01 7.576e-01 8.353e-01
## 1400 8.447e-01 9.963e-01 8.055e-01 8.855e-01
## 1401 9.819e-01 4.494e-01 6.869e-01 4.025e-01
## 1402 3.398e-01 1.751e-02 1.626e-02 1.557e-03
## 1403 1.354e-01 6.768e-01 4.367e-02 1.428e-01
## 1404 4.950e-02 9.159e-01 1.885e-02 2.483e-01
## 1405 1.923e-01 3.891e-01 3.317e-01 8.206e-01
## 1406 1.351e-01 1.218e-01 1.629e-02 1.348e-02
## 1407 5.411e-02 7.562e-01 2.229e-02 1.529e-01
## 1408 1.808e-03 1.131e-01 8.749e-03 9.834e-01
## 1409 5.711e-02 1.030e-01 2.183e-01 4.214e-01
## 1410 2.139e-02 4.456e-01 1.197e-04 1.321e-03
## 1500 7.446e-02 5.597e-02 3.787e-02 2.973e-02
## 1501 5.299e-01 2.923e-01 5.105e-01 4.248e-01
## 1502 3.126e-01 5.546e-01 2.473e-01 4.140e-01
## 1503 8.688e-03 2.147e-01 7.339e-03 1.790e-01
## 1504 9.565e-01 2.851e-01 8.944e-01 2.952e-01
## 1505      NA      NA 1.776e-01 3.004e-01
## 1506 3.769e-02 4.684e-02 2.283e-02 2.900e-02
## 1507 6.613e-01 9.908e-01 6.079e-01 7.871e-01
## 1508 1.728e-01 5.268e-01 2.056e-01 6.287e-01
## 1600 4.340e-02 3.122e-01 3.645e-02 2.536e-01
## 1601 9.445e-01 4.146e-03 6.248e-01 5.380e-03
## 1602 2.079e-01 2.134e-01 2.148e-01 2.216e-01
## 1603 2.481e-01 7.924e-01 2.231e-01 6.461e-01
## 1604 3.230e-01 2.663e-01 3.045e-01 2.528e-01
## 1605 3.603e-01 1.116e-01 3.712e-01 1.145e-01
## 1606 5.722e-03 1.122e-03 2.953e-03 6.129e-04
## 1607 3.076e-02 3.012e-01 3.330e-02 3.327e-01
## 1700 1.205e-03 4.696e-01 1.276e-03 3.227e-01
## 1701 5.606e-02 4.169e-01 1.979e-03 5.791e-02
## 1702 8.579e-01 1.115e-01 3.129e-01 5.578e-02
## 1703 4.539e-01 9.953e-01 4.378e-01 8.284e-01
## 1704 8.931e-01 5.765e-01 6.572e-01 4.938e-01
## 1705 3.064e-02 6.975e-01 2.518e-02 3.890e-01
## 1706 2.091e-01 1.390e-02 8.514e-03 8.531e-04
## 1707 5.415e-01 7.572e-01 3.939e-01 6.114e-01
## 1708 9.397e-01 2.855e-01 7.590e-01 2.663e-01
## 1709 2.457e-01 5.313e-01 1.173e-01 2.232e-01
```

```
## 1710 6.844e-01 6.738e-01 8.266e-01 7.988e-01
## 1711 8.039e-01 6.920e-01 8.937e-01 7.115e-01
## 1712 8.487e-01 6.901e-01 9.534e-01 7.177e-01
## 1800 3.666e-01 4.870e-01 1.456e-01 1.676e-01
## 1801          NA          NA          NA          NA
## 1802 6.591e-01 2.291e-01 8.357e-01 2.344e-01
## 1803 3.933e-01 7.435e-01 2.201e-01 3.104e-01
## 1804 2.676e-01 3.553e-01 1.489e-01 1.662e-01
## 1900 2.201e-01 9.042e-01 2.043e-01 6.704e-01
## 1901 4.681e-01 5.069e-02 8.474e-01 6.206e-02
## 1902 8.322e-01 3.813e-01 9.718e-01 4.027e-01
## 1903 1.100e-01 1.816e-01 2.602e-01 5.259e-01
## 1904 6.689e-01 5.300e-01 8.499e-01 6.310e-01
## 1905 2.851e-01 1.750e-01 3.498e-01 2.284e-01
## 1906 6.542e-02 9.868e-01 6.620e-02 7.731e-01
## 1907 3.961e-01 8.195e-01 3.559e-01 6.669e-01
## 1908 2.676e-02 2.215e-01 4.972e-02 4.522e-01
## 1909 3.214e-02 4.992e-01 4.792e-02 7.697e-01
## 1910 4.958e-01 4.767e-01 6.041e-01 5.812e-01
## 1911 1.446e-02 2.377e-03 3.026e-01 3.757e-02
## 1912 8.302e-02 7.551e-01 6.129e-02 4.130e-01
## 1913 5.139e-01 9.356e-01 5.164e-01 8.753e-01
## 2000 8.184e-01 3.040e-01 1.937e-01 9.790e-02
## 2001 6.800e-01 6.406e-03 3.437e-01 6.469e-03
## 2002 7.961e-01 2.804e-03 2.410e-02 2.061e-04
## 2003 9.359e-01 5.781e-03 1.202e-01 2.459e-03
## 2100          NA          NA          NA          NA
## 2101 3.670e-01 7.942e-01 5.663e-02 1.439e-01
## 2102 1.928e-01 1.108e-01 4.602e-01 1.885e-01
## 2103 4.361e-01 5.180e-01 5.123e-01 5.992e-01
## 2104 5.408e-01 2.110e-02 6.264e-01 1.219e-02
## 2105 1.558e-01 1.600e-02 7.465e-01 4.224e-02
## 2200 1.428e-01 3.910e-01 2.240e-01 7.186e-01
## 2201 2.693e-01 1.428e-01 7.470e-01 2.478e-01
## 2202 4.664e-01 6.080e-01 3.028e-01 3.733e-01
## 2203 6.569e-01 4.782e-01 8.526e-01 5.606e-01
## 2204 8.645e-01 7.503e-01 7.787e-01 6.949e-01
## 2205 4.238e-01 3.375e-01 6.469e-01 4.643e-01
## 2206 6.949e-03 5.992e-04 9.543e-03 6.977e-04
## 2207 7.285e-01 5.882e-02 7.754e-01 6.061e-02
## 2208 9.359e-01 8.033e-01 9.805e-01 8.111e-01
## 2209 1.868e-02 6.186e-01 2.950e-02 9.645e-01
## 2210 1.602e-03 4.189e-02 7.391e-04 1.776e-02
## 2211 5.359e-03 9.259e-01 5.406e-03 8.931e-01
## 2212 4.722e-01 5.384e-01 2.859e-01 2.999e-01
## 2213 1.664e-02 1.373e-01 7.644e-04 4.458e-03
## 2214 9.493e-01 5.101e-01 6.794e-01 4.343e-01
## 2215 5.356e-02 7.868e-01 2.985e-02 2.142e-01
## 2216 3.787e-02 1.560e-01 1.711e-01 8.404e-01
## 2300 3.694e-01 6.343e-03 8.381e-01 1.019e-02
```

```
## 2301 5.596e-01 6.103e-01 7.277e-01 8.766e-01
## 2302 3.088e-02 4.401e-02 1.763e-02 2.705e-02
## 2303 5.442e-01 3.597e-01 4.149e-01 2.894e-01
## 2304 7.058e-01 4.622e-03 9.566e-01 5.778e-03
## 2305 6.505e-03 9.878e-01 6.650e-03 4.974e-01
## 2306 6.042e-01 8.354e-03 2.079e-01 4.004e-03
## 2307 5.661e-01 2.718e-01 4.928e-01 2.470e-01
## 2308 1.691e-01 1.023e-01 2.115e-02 1.402e-02
## 2309 2.900e-01 3.248e-02 1.077e-01 1.501e-02
## 2310 2.396e-01 3.191e-01 1.912e-01 2.418e-01
## 2311 2.909e-02 4.288e-01 2.346e-02 2.600e-01
## 2312 1.877e-01 9.500e-01 1.885e-01 8.457e-01
## 2400 7.708e-02 6.440e-01 8.652e-02 8.154e-01
## 2401      NA      NA      NA      NA
## 2402 3.044e-01 7.044e-01 2.942e-01 6.438e-01
## 2403 1.715e-02 4.539e-03 9.879e-03 2.595e-03
## 2404 3.473e-01 9.341e-02 3.157e-01 8.620e-02
## 2405 8.154e-02 6.219e-01 6.441e-02 4.237e-01
## 2406 2.000e-01 7.952e-01 1.901e-01 6.887e-01
## 2500 5.627e-06 3.531e-01 3.441e-06 1.795e-01
## 2501 8.543e-01 6.824e-01      NA 6.664e-01
## 2502 8.484e-01 4.769e-01 7.793e-01 4.572e-01
## 2503 9.750e-01 4.222e-01 9.381e-01 4.322e-01
## 2504 8.817e-02 9.009e-01 8.842e-02 9.478e-01
## 2505 1.695e-01 5.757e-01 1.490e-01 4.796e-01
## 2506 5.463e-01 2.076e-01 3.971e-01 1.683e-01
## 2507 1.597e-01 5.149e-01 2.012e-01 7.028e-01
## 2508 1.723e-01 4.223e-01 1.611e-01 3.847e-01
## 2600 9.772e-01 7.008e-01 8.105e-01 6.601e-01
## 2601 3.734e-01 4.506e-01 9.703e-02 9.907e-02
## 2602 7.016e-01 4.779e-01 3.457e-01 2.527e-01
## 2603 2.883e-01 4.711e-01 1.255e-01 2.028e-01
## 2604 9.506e-01 2.640e-01 6.771e-01 2.336e-01
## 2605 4.894e-03 2.019e-01 6.967e-03 4.119e-01
## 2606 5.448e-02 6.690e-01 5.913e-02 9.348e-01
## 2607 5.167e-01 2.742e-01 1.170e-01 6.225e-02
## 2608 6.639e-02 5.887e-01 1.029e-01 5.642e-01
## 2609 6.874e-02 1.756e-02 7.880e-01 1.576e-01
## 2610 3.173e-01 1.796e-01 1.311e-01 7.830e-02
## 2611 5.060e-01 8.315e-01 4.030e-01 6.407e-01
## 2612 3.184e-01 3.219e-01 4.395e-01 5.175e-01
## 2613 6.190e-01 6.839e-01 6.918e-01 8.030e-01
## 2614 9.257e-02 3.718e-01 1.586e-01 7.777e-01
## 2700 1.554e-09 1.652e-03 4.035e-07 3.177e-01
## 2701 2.762e-05 3.843e-01 4.455e-05 8.590e-01
## 2702 8.477e-02 6.449e-01 5.890e-02 3.299e-01
## 2703 2.327e-03 1.952e-01 1.777e-03 1.263e-01
## 2704 8.923e-01 1.291e-01 9.837e-01 1.279e-01
## 2705 4.646e-07 8.127e-04 1.489e-08 2.262e-05
## 2706 1.039e-01 7.469e-02 4.924e-02 3.307e-02
```

```
## 2707 2.303e-02 6.911e-03 2.221e-01 7.164e-02
## 2708 6.376e-01 8.800e-01 6.497e-01 9.228e-01
## 2709      NA      NA      NA      NA
## 2710 3.515e-01 9.358e-01 3.219e-01 8.341e-01
## 2711 2.215e-03 3.878e-02 9.569e-04 1.508e-02
## 2712 3.011e-01 9.233e-02 3.491e-01 1.055e-01
## 2713 4.065e-03 6.538e-01 4.655e-03 9.363e-01
## 2714 2.780e-03 7.894e-01 1.586e-03 2.716e-01
## 2715 6.293e-01 2.679e-01 5.625e-01 2.506e-01
## 2716 1.512e-02 2.256e-01 9.362e-03 1.220e-01
## 2717 2.217e-01 9.798e-01 2.126e-01 7.358e-01
## 2718 1.723e-03 8.578e-01 4.023e-04 1.068e-01
## 2719 8.255e-03 3.423e-01 1.396e-02 9.539e-01
## 2720 3.134e-01 8.779e-03 2.663e-01 8.011e-03
## 2721 8.582e-01 9.425e-01 8.626e-01 9.497e-01
## 2722 4.756e-01 6.838e-01 4.693e-01 6.709e-01
## 2723 4.019e-02 1.011e-01 3.448e-02 8.409e-02
## 2724 8.476e-02 1.890e-01 9.140e-02 2.029e-01
## 2725 6.488e-01 6.413e-01 6.893e-01 6.765e-01
## 2726 2.722e-01 7.658e-01 2.730e-01 7.714e-01
## 2727 5.921e-01 8.483e-01 5.839e-01 8.210e-01
## 2728 3.628e-07 3.037e-03 5.408e-09 2.879e-05
## 2729 1.440e-08 3.902e-04 3.031e-12 3.102e-08
## 2730 4.684e-01 5.667e-01 5.246e-01 6.506e-01
## 2731 1.464e-01 1.454e-01 2.252e-01 2.178e-01
## 2732 5.347e-01 3.946e-01 3.771e-01 2.962e-01
## 2733 1.320e-01 7.001e-01 5.787e-02 2.184e-01
## 2734 7.902e-03 3.368e-01 1.523e-02 7.582e-01
## 2735 1.815e-06 3.916e-05 4.586e-09 6.724e-08
## 2736 4.216e-02 7.490e-03 1.447e-01 2.316e-02
## 2737 1.741e-03 6.925e-01 1.869e-03 9.812e-01
## 2738 5.117e-15 2.104e-02 4.735e-16 5.419e-04
## 2739 1.372e-06 5.445e-01 1.879e-06 4.856e-01
## 2740 2.352e-05 1.979e-01 4.206e-06 1.858e-02
## 2741 8.831e-03 9.516e-04 8.863e-04 1.270e-04
## 2742 1.931e-02 4.863e-01 2.889e-02 8.955e-01
## 2743 2.672e-01 3.112e-01 3.540e-01 4.274e-01
## 2744      NA      NA      NA      NA
## 2745 2.099e-01 2.534e-01 1.696e-01 1.998e-01
## 2746 2.107e-06 1.425e-03 1.826e-07 1.124e-04
## 2747 6.756e-02 1.066e-03 6.411e-02 1.241e-03
## 2748 9.116e-01 9.927e-01 9.138e-01 9.648e-01
## 2800 1.244e-01 2.177e-01 1.804e-01 3.376e-01
## 2801 4.211e-01 3.752e-01 4.687e-01 4.082e-01
## 2802 9.297e-01 1.982e-01 8.544e-01 1.977e-01
## 2803 1.673e-01 4.715e-03 1.279e-01 3.704e-03
## 2804 9.031e-01 1.216e-01 7.710e-01 1.172e-01
## 2805 8.411e-01 1.642e-01 6.316e-01 1.533e-01
## 2806 4.997e-02 8.332e-02 2.574e-02 4.162e-02
## 2807 7.644e-01 7.197e-01 7.792e-01 7.031e-01
```



```
## 2808 1.826e-05 4.721e-01 5.774e-06 8.256e-02
## 2809 8.557e-01 7.551e-01 8.170e-01 7.369e-01
## 2900 1.607e-01 6.834e-04 5.455e-03 3.687e-05
## 2901          NA          NA 1.470e-01 2.301e-01
## 2902 1.705e-01 7.099e-06 1.269e-02 1.054e-06
## 2903          NA          NA          NA          NA
## 2904          NA          NA          NA          NA
## 2905 9.075e-01 1.322e-01 6.118e-01 1.189e-01
## 2906 1.434e-01 9.500e-04 7.156e-01 3.746e-03
## 2907 1.772e-02 3.652e-01 3.782e-02 8.461e-01
## 2908 1.951e-01 1.055e-04 9.141e-01 8.676e-04
## 2909 3.696e-01 5.627e-01 3.036e-01 4.364e-01
## 2910 3.452e-01 4.258e-01 5.354e-01 9.048e-01
## 2911 9.555e-01 4.960e-01 6.234e-01 4.191e-01
## 2912 5.374e-02 5.830e-01 6.482e-02 8.778e-01
## 2913 6.864e-01 1.428e-01 7.200e-01 1.517e-01
## 2914 5.897e-01 4.097e-01 9.720e-01 3.985e-01
## 2915          NA          NA          NA          NA
## 2916 4.338e-01 1.863e-03 8.748e-01 2.374e-03
## 2917 5.089e-02 5.109e-01 4.597e-02 4.679e-01
## 2918          NA          NA          NA          NA
## 2919 9.428e-01 3.253e-01 7.528e-01 2.892e-01
## 2920          NA          NA          NA          NA
## 2921 7.624e-01 2.121e-02 8.934e-01 2.484e-02
## 2922 9.962e-02 5.563e-01 1.711e-02 1.175e-01
## 2923 8.466e-02 1.960e-01 2.710e-01 3.204e-01
## 3000 7.001e-01 5.721e-03 4.566e-01 4.225e-03
## 3001          NA          NA          NA          NA
## 3002 4.115e-01 1.509e-01 4.182e-01 1.521e-01
## 3003 8.283e-01 3.498e-03 7.305e-01 3.814e-03
## 3004 2.252e-01 7.417e-02 2.799e-01 8.948e-02
## 3005 4.726e-01 8.720e-02 5.604e-01 9.988e-02
## 3100 6.296e-01 3.028e-01 4.701e-01 2.578e-01
## 3101 7.301e-01 5.262e-01 6.479e-01 4.805e-01
## 3102 6.621e-01 4.431e-02 8.424e-01 4.360e-02
## 3103 5.401e-01 5.748e-01 6.673e-01 7.095e-01
## 3104 2.118e-02 5.936e-01 2.445e-02 7.777e-01
## 3105 5.809e-01 9.413e-01 5.871e-01 9.756e-01
## 3106 9.741e-01 5.573e-01 9.173e-01 5.695e-01
## 3107 8.506e-01 5.101e-01 8.344e-01 5.062e-01
## 3108 4.154e-01 9.063e-01 4.323e-01 9.857e-01
## 3109 7.052e-02 5.340e-01 3.758e-02 2.858e-01
## 3110 3.072e-01 9.116e-01 3.151e-01 9.566e-01
## 3200 3.505e-01 1.197e-04 3.857e-02 2.898e-05
## 3201 2.609e-03 5.623e-01 1.763e-03 2.638e-01
## 3202 1.646e-02 4.136e-01 5.678e-03 9.098e-02
## 3203 2.331e-01 7.881e-01 2.445e-01 9.228e-01
## 3204 3.854e-01 5.717e-03 8.851e-01 9.919e-03
## 3205 6.389e-03 3.642e-01 3.131e-03 1.279e-01
## 3206 6.773e-02 7.469e-02 3.697e-02 3.997e-02
```

```
## 3207 1.054e-03 3.397e-02 1.127e-02 4.473e-01
## 3300 5.585e-01 7.308e-01 2.688e-01 3.154e-01
## 3301 5.766e-01 2.005e-02 2.165e-01 1.111e-02
## 3302 4.728e-01 3.316e-01 1.004e-01 7.909e-02
## 3303 1.069e-02 3.210e-01 8.075e-03 2.200e-01
## 3304 9.201e-02 9.300e-01 4.338e-02 2.268e-01
## 3305 6.973e-03 8.610e-01 3.906e-04 1.579e-02
## 3306 1.006e-04 4.426e-01 1.817e-04 3.916e-01
## 3307 2.151e-03 1.069e-01 3.973e-04 9.999e-03
## 3308 6.691e-02 9.660e-02 3.366e-01 5.877e-01
## 3309 7.413e-01 5.297e-01 8.509e-01 5.598e-01
## 3310 2.905e-04 8.908e-02 7.889e-04 1.999e-01
## 3311 8.876e-01 7.190e-03 8.138e-02 2.036e-03
## 3312 3.548e-01 2.626e-01 1.081e-02 8.724e-03
## 3313 1.711e-01 3.892e-01 7.284e-02 1.427e-01
## 3314 9.233e-01 9.936e-01 8.844e-01 9.129e-01
## 3315 3.779e-02 2.051e-01 8.378e-02 5.966e-01
## 3316 9.422e-01 3.534e-01 2.062e-01 1.179e-01
## 3317 1.819e-01 4.004e-01 1.192e-02 2.123e-02
## 3318 1.521e-01 6.941e-02 8.520e-01 3.313e-01
## 3319 6.198e-01 1.168e-01 8.668e-01 1.703e-01
## 3320 1.411e-01 9.173e-01 7.120e-03 2.472e-02
## 3321 4.968e-01 1.293e-02 3.917e-01 1.036e-02
## 3322 2.517e-01 7.795e-02 9.059e-01 2.013e-01
## 3400 5.578e-02 1.382e-01 4.617e-02 1.081e-01
## 3401      NA      NA      NA      NA
## 3402      NA      NA      NA      NA
## 3403 2.522e-02 4.503e-02 2.822e-02 5.549e-02
## 3404 6.929e-01 4.912e-01 7.265e-01 5.153e-01
## 3500 2.249e-01 5.217e-01 3.347e-01 9.273e-01
## 3501      NA      NA      NA      NA
## 3502      NA      NA      NA      NA
## 3503      NA      NA      NA      NA
## 3504 5.861e-01 2.627e-01 9.797e-01 3.396e-01
## 3505      NA      NA      NA      NA
## 3506      NA      NA      NA      NA
## 3600 2.878e-02 3.002e-01 5.998e-02 9.259e-01
## 3601 9.321e-02 6.078e-01 1.214e-01 6.172e-01
## 3602 6.921e-01 1.144e-01 1.509e-01 2.027e-02
## 3603 1.118e-01 6.177e-01 1.237e-01 7.620e-01
## 3604 4.923e-01 3.403e-01 1.996e-01 1.314e-01
## 3605 6.973e-01 4.292e-01 9.538e-01 4.577e-01
## 3606      NA      NA      NA      NA
## 3607 6.089e-01 1.049e-01 2.431e-01 5.530e-02
## 3608      NA      NA      NA      NA
## 3609 3.448e-04 7.699e-01 4.686e-04 4.054e-01
## 3610 3.626e-03 8.236e-01 1.725e-03 1.980e-01
## 3611 6.005e-02 2.625e-01 1.985e-02 7.893e-02
## 3612 4.736e-01 1.451e-01 1.918e-01 7.216e-02
## 3613 5.897e-01 7.394e-01 5.709e-01 6.965e-01
```

```
## 3614 5.787e-01 8.322e-01 5.447e-01 7.290e-01
## 3615          NA          NA          NA          NA
## 3616 3.253e-01 7.240e-02 5.358e-01 1.101e-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 "" "" "****" "****" "****" "****" "****" "" "" "" ""
```

[illegible]


```

## 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          259         120         73
## 1100          404         208        148

```

## 1101	39	16	8
## 1102	181	55	29
## 1103	313	169	97
## 1104	350	145	79
## 1105	764	332	204
## 1106	232	144	91
## 1107	88	33	18
## 1108	30	8	6
## 1109	76	33	9
## 1110	233	82	55
## 1111	135	38	24
## 1200	34	12	15
## 1201	387	230	201
## 1202	410	206	189
## 1203	210	130	115
## 1204	72	44	33
## 1205	2	1	1
## 1206	26	15	15
## 1207	100	52	45
## 1208	186	108	108
## 1209	22	13	14
## 1210	45	28	23
## 1211	210	70	75
## 1212	106	41	44
## 1213	182	116	111
## 1300	470	250	175
## 1301	5	2	1
## 1302	30	13	13
## 1303	283	116	69
## 1304	115	44	37
## 1305	132	48	32
## 1306	142	87	44
## 1307	185	91	51
## 1308	62	25	17
## 1309	71	49	26
## 1310	110	72	54
## 1311	322	157	99
## 1312	326	132	97
## 1313	127	56	35
## 1314	232	101	75
## 1315	35	15	8
## 1400	100	39	32
## 1401	56	23	18
## 1402	93	30	24
## 1403	181	68	70
## 1404	37	5	11
## 1405	81	35	31
## 1406	129	66	64
## 1407	121	54	52
## 1408	212	80	87

## 1409	151	93	94
## 1410	43	16	13
## 1500	145	36	14
## 1501	15	4	2
## 1502	66	17	13
## 1503	62	19	13
## 1504	5	3	1
## 1505	18	5	3
## 1506	14	4	4
## 1507	26	7	1
## 1508	31	10	6
## 1600	355	103	60
## 1601	22	8	6
## 1602	87	36	16
## 1603	25	9	6
## 1604	61	18	14
## 1605	138	47	23
## 1606	158	49	30
## 1607	52	22	13
## 1700	54	17	13
## 1701	10	2	2
## 1702	9	2	3
## 1703	36	6	6
## 1704	23	7	8
## 1705	41	10	11
## 1706	98	26	19
## 1707	19	4	2
## 1708	26	3	5
## 1709	63	22	23
## 1710	84	24	19
## 1711	37	4	4
## 1712	96	17	23
## 1800	31	15	14
## 1801	1	1	0
## 1802	43	8	10
## 1803	54	14	8
## 1804	43	5	5
## 1900	66	16	15
## 1901	53	18	7
## 1902	110	53	31
## 1903	25	10	10
## 1904	154	73	55
## 1905	26	10	7
## 1906	103	30	14
## 1907	118	43	27
## 1908	56	18	6
## 1909	134	21	19
## 1910	169	77	39
## 1911	51	28	6
## 1912	62	15	4

## 1913	17	5	5
## 2000	65	13	17
## 2001	67	21	16
## 2002	381	110	92
## 2003	140	31	23
## 2100	56	14	12
## 2101	16	2	3
## 2102	85	16	14
## 2103	58	12	14
## 2104	31	11	10
## 2105	182	54	41
## 2200	42	17	8
## 2201	36	18	14
## 2202	36	4	1
## 2203	30	12	8
## 2204	133	40	26
## 2205	221	45	31
## 2206	12	1	0
## 2207	45	3	2
## 2208	99	11	9
## 2209	129	29	22
## 2210	193	29	21
## 2211	126	18	8
## 2212	49	15	6
## 2213	89	31	26
## 2214	14	3	4
## 2215	152	32	23
## 2216	44	16	12
## 2300	195	67	53
## 2301	95	54	45
## 2302	48	17	19
## 2303	487	204	136
## 2304	179	68	40
## 2305	155	49	38
## 2306	95	41	26
## 2307	161	99	63
## 2308	379	179	129
## 2309	202	94	61
## 2310	176	67	51
## 2311	117	43	34
## 2312	189	64	35
## 2400	83	41	30
## 2401	2	1	1
## 2402	50	25	12
## 2403	176	105	63
## 2404	148	90	43
## 2405	67	32	16
## 2406	73	49	24
## 2500	273	58	55
## 2501	11	6	5

## 2502	81	20	13
## 2503	47	11	8
## 2504	83	17	14
## 2505	65	17	15
## 2506	70	17	13
## 2507	62	20	8
## 2508	90	16	18
## 2600	90	20	15
## 2601	18	5	3
## 2602	29	2	3
## 2603	26	3	1
## 2604	124	15	15
## 2605	29	4	1
## 2606	35	4	4
## 2607	19	2	4
## 2608	13	1	2
## 2609	2	0	0
## 2610	29	3	3
## 2611	74	16	15
## 2612	7	0	0
## 2613	104	16	12
## 2614	27	1	3
## 2700	477	273	205
## 2701	324	217	152
## 2702	37	12	14
## 2703	107	55	43
## 2704	15	3	5
## 2705	294	93	63
## 2706	92	46	29
## 2707	61	29	27
## 2708	98	56	34
## 2709	2	2	0
## 2710	9	8	5
## 2711	138	74	45
## 2712	165	105	68
## 2713	106	62	53
## 2714	55	37	26
## 2715	93	28	21
## 2716	82	52	30
## 2717	142	92	80
## 2718	137	84	62
## 2719	395	269	227
## 2720	76	48	25
## 2721	40	11	9
## 2722	35	17	13
## 2723	115	76	49
## 2724	166	97	51
## 2725	285	154	92
## 2726	115	58	36
## 2727	61	29	18

## 2728	351	169	135
## 2729	290	231	195
## 2730	487	212	122
## 2731	123	59	34
## 2732	419	162	114
## 2733	81	47	34
## 2734	230	123	86
## 2735	389	252	201
## 2736	230	138	101
## 2737	180	61	65
## 2738	685	424	311
## 2739	907	620	485
## 2740	155	70	53
## 2741	364	82	56
## 2742	135	103	84
## 2743	62	51	37
## 2744	0	0	0
## 2745	35	17	10
## 2746	437	142	76
## 2747	29	12	6
## 2748	70	25	17
## 2800	198	93	67
## 2801	25	18	13
## 2802	116	87	47
## 2803	95	58	40
## 2804	105	57	35
## 2805	110	71	48
## 2806	31	21	15
## 2807	18	14	8
## 2808	183	88	68
## 2809	75	38	20
## 2900	307	242	227
## 2901	15	14	6
## 2902	56	45	28
## 2903	3	2	2
## 2904	6	4	2
## 2905	78	58	47
## 2906	27	19	19
## 2907	58	36	32
## 2908	7	6	3
## 2909	49	38	33
## 2910	39	21	17
## 2911	15	12	11
## 2912	33	30	23
## 2913	73	71	66
## 2914	13	13	0
## 2915	1	0	0
## 2916	274	229	175
## 2917	10	9	6
## 2918	0	0	0

## 2919	31	28	27
## 2920	4	3	2
## 2921	94	56	57
## 2922	18	18	0
## 2923	13	11	9
## 3000	26	18	9
## 3001	6	4	3
## 3002	98	30	15
## 3003	81	43	35
## 3004	194	102	75
## 3005	88	48	32
## 3100	154	35	19
## 3101	26	8	2
## 3102	25	8	7
## 3103	46	8	5
## 3104	194	30	23
## 3105	61	12	13
## 3106	20	3	1
## 3107	99	19	9
## 3108	28	6	7
## 3109	35	3	3
## 3110	19	3	4
## 3200	281	157	135
## 3201	40	24	21
## 3202	204	121	99
## 3203	309	186	162
## 3204	394	277	230
## 3205	229	125	101
## 3206	97	59	50
## 3207	239	157	146
## 3300	177	80	79
## 3301	246	162	153
## 3302	68	40	32
## 3303	157	96	78
## 3304	1248	825	757
## 3305	559	289	278
## 3306	360	257	216
## 3307	63	26	24
## 3308	454	245	221
## 3309	88	52	45
## 3310	243	152	142
## 3311	69	24	21
## 3312	960	537	482
## 3313	104	34	25
## 3314	216	127	117
## 3315	216	131	128
## 3316	507	276	274
## 3317	58	32	37
## 3318	109	86	78
## 3319	68	42	46

## 3320	298	110	107
## 3321	115	64	56
## 3322	126	65	58
## 3400	141	100	58
## 3401	10	5	2
## 3402	5	2	2
## 3403	36	24	11
## 3404	8	4	2
## 3500	23	9	4
## 3501	0	0	0
## 3502	0	0	0
## 3503	0	0	0
## 3504	13	3	1
## 3505	3	2	1
## 3506	2	2	0
## 3600	9	7	9
## 3601	18	12	12
## 3602	16	6	5
## 3603	6	4	3
## 3604	27	18	16
## 3605	34	13	11
## 3606	0	0	0
## 3607	9	3	1
## 3608	1	1	0
## 3609	54	46	44
## 3610	30	16	11
## 3611	80	58	49
## 3612	361	192	145
## 3613	3	1	1
## 3614	53	16	17
## 3615	0	0	0
## 3616	106	90	72

```

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="")
```